



UNIVERSIDAD
DE PIURA

FACULTAD DE INGENIERÍA

**Implementación de la cadena de suministro de información
climática a las estaciones meteorológicas Niño5 de la
Estación Científica de la Universidad de Piura**

Tesis para optar el Título de
Ingeniero Industrial y de Sistemas

Jefferson Stefan Llacsahuanga Pusma

Asesor(es):

Dr. Ing. Rodolfo Rodríguez Arisméndiz

Dr. Ing. Mario José Quinde Li Say Tan

Piura, marzo de 2023

NOMBRE DEL TRABAJO

Tesis_IIS_LlacsahuangaJefferson (2).docx

RECuento DE PALABRAS

13915 Words

RECuento DE CARACTERES

75534 Characters

RECuento DE PÁGINAS

80 Pages

TAMAÑO DEL ARCHIVO

7.2MB

FECHA DE ENTREGA

Mar 14, 2023 9:03 AM GMT-5

FECHA DEL INFORME

Mar 14, 2023 9:04 AM GMT-5**● 10% de similitud general**

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos

- 8% Base de datos de Internet
- 3% Base de datos de publicaciones
- Base de datos de Crossref
- Base de datos de contenido publicado de Crossref
- 6% Base de datos de trabajos entregados

Dedicatoria

A Dios y a la Virgen, por darme la bendición de seguir adelante.

A mis padres, quienes son los que me dan la motivación y oportunidad para poder seguir desarrollándome en mi rubro profesional.



Resumen

En este trabajo se ha implementado una cadena de suministro de información climática de la cadena estaciones meteorológicas Niño 5 (NIÑO5) que administra la UDEP. En esta implementación se han establecido etapas en base al ciclo de vida de los datos utilizando los productos del software Tableau.

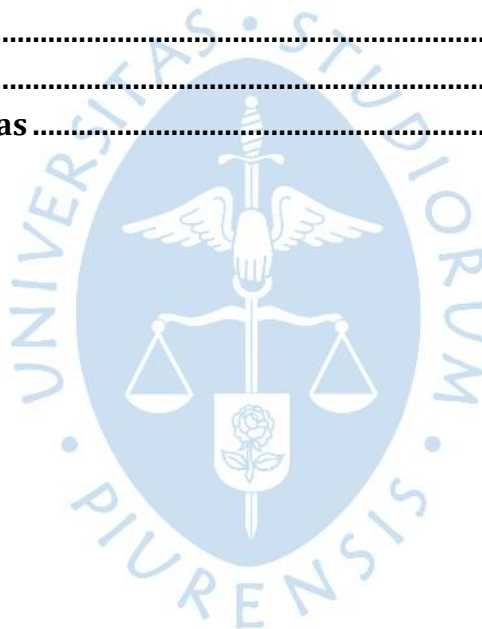
NINO5 es un conjunto de diez estaciones instaladas en lugares cercanos a la latitud de 5°S, entre las regiones de Piura y Cajamarca y son administradas desde la estación científica Ramón Mugica de UDEP. El análisis de los parámetros meteorológicos generados por esta cadena puede tener diversas aplicaciones, como contribuir en la decisión de acciones preventivas ante fenómenos climáticos como El Niño, determinar potencial de energías renovables, productividad en el sector agrario, entre otros. Para obtener estas aplicaciones es necesario tener procesos eficientes de preparación de datos, para realizar razonamientos descriptivos y predictivos.

Se hace énfasis sobre la influencia del *Business Analytics* y su aplicación en la correcta toma de decisiones en distintos rubros. Resaltando que el rubro investigación meteorológica y climática no es ajena a estas implementaciones. Se detalla diversos retos que presentan los procesos de preparación de datos de las estaciones meteorológicas ante el comportamiento de datos atípicos, faltantes, y formatos distintos. Además, se detallan criterios y tareas necesarias con sus funciones respectivas para poder corregir errores de datos. También, se definen los criterios para la elaboración de correctas visualizaciones. Finalmente, se propone una lista de usuarios y pasos necesarios para la distribución de la información trabajada. Es importante resaltar, que la propuesta de implementación de la cadena de suministro de información será totalmente adaptable al estudio de nuevos parámetros y estaciones meteorológicas que se deseen examinar.

Tabla de contenido

Introducción	15
Capítulo 1 Antecedentes.....	17
1.1 Business analytics.....	17
1.1.1 Definición.....	17
1.1.2 Ciclo de vida de datos	18
1.1.3 Cadena de suministro de información	21
1.2 Fenómeno el Niño y estaciones meteorológicas en Piura.....	22
1.2.1 Fenómeno El Niño	22
1.2.2 Cadena de estaciones meteorológicas Niño 5	24
1.3 Cadena de suministro de información en la cadena NIÑO5.....	28
1.3.1 Estado actual	28
1.3.2 Problemática.....	31
1.3.3 Propuesta de solución	33
Capítulo 2 Metodología	35
2.1 Obtención y descripción de fuente de datos.....	35
2.2 Software.....	39
2.2.1 Tableau Prep Builder	39
2.2.2 Tabpy	39
2.2.3 Tableau Desktop	39
2.2.4 Tableau Online.....	40
2.3 Preparación de datos	40
2.3.1 Limpieza.....	40
2.3.2 Ordenamiento.....	42
2.4 Análisis y distribución de información	42

2.4.1	Diseño e implementación de visualizaciones	43
2.4.2	Validación	44
Capítulo 3 Diseño e implementación de la cadena de suministro de información		47
3.1	Fuente de datos	47
3.2	Preparación de datos	48
3.2.1	Ordenamiento.....	48
3.2.2	Limpieza de datos.....	53
3.2.3	Integración	59
3.3	Visualizaciones.....	61
3.4	Receptores de información	71
3.5	Implementación	74
Conclusiones		77
Glosario.....		81
Referencias bibliográficas		83



Lista de tablas

Tabla 1. Parámetros océano – atmosféricos del FEN.....	24
Tabla 2. Ubicación de las estaciones meteorológicas que conforman la cadena NIÑO5	25
Tabla 3. Aplicaciones de estudio con datos registrados por la cadena NIÑO5	27
Tabla 4. Parámetros medidos en la cadena NIÑO5	27
Tabla 5. Parámetros registrados por cada estación meteorológica	30
Tabla 6. Identificación de parámetros	36
Tabla 7. Valores sexagesimales en dirección del viento.	37
Tabla 8. Resumen de las fuentes de datos con sus respectivos registros	38
Tabla 9. Formatos de los archivos de datos	39
Tabla 10. Nombre y formato de los valores de los parámetros	40
Tabla 11. Rango de valores de datos atípicos	41
Tabla 12. Criterios de agregación	42
Tabla 13. Criterios de visualizaciones	43
Tabla 14. Relación de <i>stakeholders</i>	72
Tabla 15. Roles de usuarios registrados.....	73
Tabla 16. Presentación de visualizaciones	76



Lista de figuras

Figura 1. Proceso de Business Analytics organizacional.	18
Figura 2. Analítica en procesos de calidad.	20
Figura 3. Mecanismo de interacción Océano Atmosfera en el Océano Pacífico Ecuatorial	23
Figura 4. Variaciones de temperatura durante el Niño	23
Figura 5. Mapa de ubicaciones de estaciones meteorológicas de la cadena NIÑO5.	26
Figura 6. Matriz de datos radiación solar de la estación Colán	31
Figura 7. Ejemplo de datos de humedad relativa de la estación de Colan, año 2013.	32
Figura 8. Gráfico de variación mensual de la humedad relativa de la estación Colán, año 2013	32
Figura 9. Matriz de datos anómalos y nulos (null) para la estación Colán.....	33
Figura 10. Diagrama de flujo cadena de suministro de información	45
Figura 11. Procedimientos de la cadena de suministro de información	47
Figura 12. Cambio de estructuras	48
Figura 13. Configuración inicial.....	48
Figura 14. Integración por comodín. A) Integración por años. B) Integración por comodín.	49
Figura 15. Cambios de formato en bloques de limpieza.....	49
Figura 16. Unión de columnas en el campo de fecha y hora. A) Formato inicial. B) Campo calculado. C) Pasos empleados. D) Resultado final de unión de campos.	50
Figura 17. Formato de fechas distintas	51
Figura 18. Matriz de datos repetidos en la Estación Colán	51
Figura 19. Matriz de datos repetidos en la Estación Chávez.....	52
Figura 20. Eliminación de duplicados con paso "Agregar"	52

Figura 21. Paso de "Salida" en flujo UDEP	53
Figura 22. Anomalías en la estación Colán. A) Anomalías Radiación Solar. B) Anomalías velocidad de viento. C) Anomalías Dirección del viento. D) Anomalías Presión atmosférica.	54
Figura 23. Anomalías en la estación Chávez. A) Anomalías temperatura. B) Anomalías Radiación Solar. C) Anomalías Lluvia acumulada.....	54
Figura 24. Funciones para limpieza de datos temperatura.....	55
Figura 25. Funciones para limpieza de temperatura (Huancabamba y San Ignacio)	55
Figura 26. Función para limpieza de los datos de Humedad.....	55
Figura 27. Función de limpieza para la dirección del viento	56
Figura 28. Función de limpieza para la velocidad del viento.....	56
Figura 29. Funciones para limpieza de parámetro de radiación solar.....	56
Figura 30. Funciones para limpieza de presión atmosférica.	57
Figura 31. Algoritmo Python empleado en datos nulos.....	58
Figura 32. Configuración Tabpy	58
Figura 33. Función para separar las fecha y hora	59
Figura 34. Pasos de agregación.....	59
Figura 35. Pasos de "Unión" de columnas.....	60
Figura 36. Campos de ubicación	60
Figura 37. Flujo final de preparación	61
Figura 38. Ejemplo de climograma.....	62
Figura 39. Gráficos de Visualización de velocidad de viento, humedad relativa y presión atmosférica.	63
Figura 40. Visualizaciones de radiación y energía solar.	64
Figura 41. Visualizaciones para frecuencia y distribución del viento.	65
Figura 42. Visualizaciones de temperatura, presión, humedad y velocidad del viento anual.	66
Figura 43. Visualización de lluvia acumulada anual.	67
Figura 44. Energía acumulada anualmente	67
Figura 45. Distribución y frecuencia de vientos anuales.....	68

Figura 46. Algoritmos para energía solar promedio y máxima.	69
Figura 47. Algoritmo para agrupación de dirección del viento.	69
Figura 48. Filtros de dashboard.	70
Figura 49. Dashboard mensual Niño 5	70
Figura 50. Dashboard anual Niño5.....	71
Figura 51. Pasos para subir flujos a Tableau Online	74
Figura 52. Dashboard Tableau Online	75
Figura 53. Permisos respecto a contenido Tableau Online.	75





Introducción

Actualmente, el mundo organizacional explota los datos en incontables modelos analíticos con el fin de tomar decisiones y acciones concretas en diferentes ámbitos. Por ejemplo, rubros comerciales, sociales, científicos, entre otros. Dentro de este panorama se encuentra el BA el cual asocia distintas disciplinas para obtener procesos descriptivos, predictivos y prescriptivos permitiendo tener una metodología inteligente en la toma de decisiones usando grandes y diversos volúmenes de datos. Por tal motivo, un gran número de organizaciones apuestan por la implementación y mejora del suministro de información para poder tener un desarrollo eficiente y ordenado en el procesamiento de información con el fin de distribuir la información en el tiempo y al responsable correcto.

El rubro de la investigación no es ajeno a la aplicación de estas metodologías de toma de decisiones inteligentes y suministros de información eficientes. En la actualidad, la estación científica Ramón Mugica de la UDEP administra una cadena estaciones meteorológicas instaladas en la región Piura y Cajamarca denominadas NIÑO5. Con la finalidad de monitorear una serie de parámetros meteorológicos para brindar diversas aplicaciones complementarias al ámbito de prevención de fenómenos climáticos, comportamiento del sector productivo agrícola, energético, salud, etc. Para brindar estas aplicaciones se emplean modelos descriptivos y predictivos para estudiar los parámetros meteorológicos relacionados al caso de estudio planteado. Sin embargo, para poder trabajar con este razonamiento se necesita poseer un suministro de información eficiente que permita integrar, preparar, visualizar y distribuir un gran volumen de datos provenientes de las estaciones meteorológicas.

La implementación de una cadena suministro de información en la cadena NIÑO5 permitirá emplear procesos de manejo de información de manera ordenada basándose en el ciclo de vida de los datos. Dejando de esta manera los métodos tradicionales de manejo de datos y optar por procesos eficientes que permitan compartir la información correcta en base a los requerimientos del usuario en el tiempo adecuado. Y finalmente, sea adaptable para aplicaciones futuras de predicción, monitoreo en vivo de parámetros, entre otras.



Capítulo 1

Antecedentes

1.1 Business analytics

En la actualidad, las organizaciones apuestan a sobresalir en distintas áreas con el objetivo de tener una ventaja competitiva dentro de sus rubros. Davenport (2013) menciona que existe un interés en lograr un continuo crecimiento basándose en el análisis y la ciencia de los datos. Por tal razón, las organizaciones apuestan al uso de distintas fuentes de datos para generar valor dentro de los rubros en los que se encuentran. Hindle, Kunc, Mortensen, Oztekin, & Vidgen (2019) agrega una perspectiva interesante al referirse que las propuestas de valor están basadas en métodos analíticos orientadas a predecir, por ejemplo, comportamientos de consumidores o probabilidades para definir condiciones médicas de los pacientes.

1.1.1 Definición

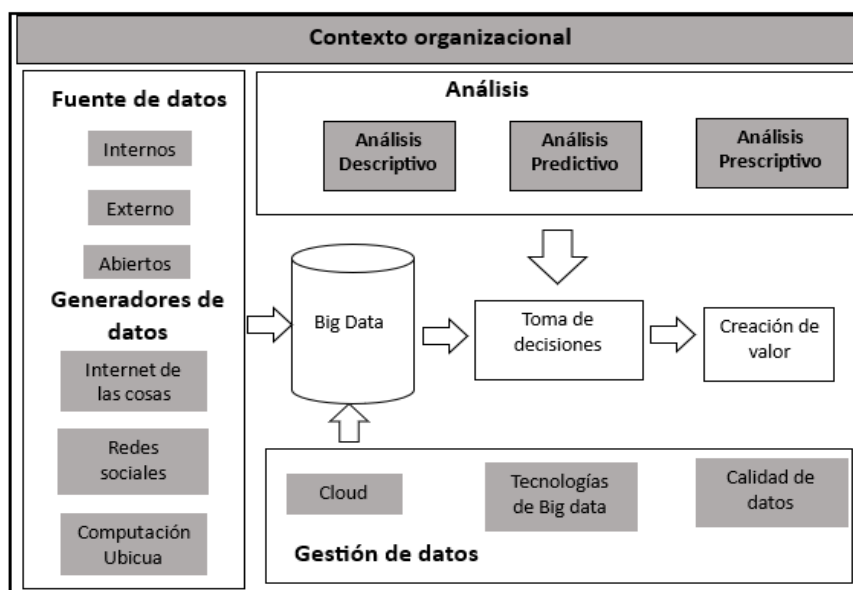
Omar, Minoufekar, & Plapper (2019) define el *business analytics* (BA) como un análisis cuyo fin es la resolución de problemas, en base a una serie de evidencias obtenidas mediante la colaboración del trabajo humano y sistemas dentro de la organización. Enfatiza reiteradamente que la data es la evidencia para poder plantear soluciones en base a lo requerido. Este concepto se complementa con la definición de Davenport (2007), enfatizando la analítica como el uso de datos en análisis estadísticos y cuantitativos, modelos de predicción y otras acciones con la finalidad de tomar decisiones que generen una ventaja competitiva

Holsapple, Lee-Post, & Pakth (2014) recalca el dominio que tiene el BA en las distintas disciplinas de las organizaciones, como en marketing, recursos humanos, negocios estrategia, operaciones, sistemas de cadena de suministro, sistemas de información y finanzas, y dentro de estas se encuentran subdimensiones orientadas a las soluciones basadas en innovación y competitividad.

La aplicación del BA resalta la importancia de su implementación en los distintos rubros de las organizaciones. Por ejemplo, en la gestión de cadenas de suministro sobresalen distintas aplicaciones. Omar, Minoufekar & Plapper (2019) da a conocer como la correcta información puede aumentar el rendimiento de esta. Enfatizando que, BA ayuda a tener una

mejor relación entre todos los agentes para gestionar entregas correctas entre consumidores y proveedores, logrando así la ventaja competitiva. Además, añade una relación entre la previsión de demanda y la eficiencia en la gestión de recursos, señalando a esta primera como la responsable de respaldar la toma de decisiones estratégicas como expansión de capacidad, innovación en tecnología, etc. Siguiendo en la línea de gestión de cadena de suministros se encuentran más procesos involucrados con el BA, como lo es la producción, calidad e inventarios. Sin embargo, es importante dar mención que todo esto es parte del proceso organizacional de BA como lo determina Hindle, Kunc, Mortensen, Oztekin, & Vidgen (2019) en el marco de la Figura 1.

Figura 1. Proceso de Business Analytics



Nota. Adaptado de “Business analytics: Defining the field and identifying a research” Hindle, Kunc, Mortensen, Oztekin, & Vidgen (2019).

1.1.2 Ciclo de vida de datos

La orientación del uso correcto de los datos es hacia la obtención de la ventaja competitiva. Esto es posible siguiendo el ciclo de vida de datos. Perera , Zaslavsky, Christen, & Georgakopoulos (2014) determina 4 fases en este ciclo: adquisición, preparación, razonamiento y distribución.

A. Adquisición.

Uno de los principales retos iniciales para las organizaciones es la adquisición de los datos, ya que, las organizaciones deben buscar las fuentes correctas para poder entablar la toma de decisiones de forma adecuada. Nagle & Sammon (2016) menciona como una problemática, la mala calidad de datos, debido a que algunas organizaciones todavía apuestan a tener solo una fuente de datos como lo son hojas de cálculos en donde se presentan

inexactitudes, duplicados y errados, ocasionando dificultades en sus recopilaciones. Por otra parte, Holsapple, Lee-Post, & Pakth (2014) da a conocer que los datos pueden encontrarse de forma estructurada, semiestructurados y no estructurados.

B. Preparación.

La preparación de datos se basa en principios de integración, limpieza y ordenamiento. Debido a que en la adquisición se obtienen fuentes de datos heterogéneos, se tiende a originar información inusual e irrelevante para el posterior razonamiento de estos. Resaltando la importancia de este procedimiento, Omar, Minoufekar, & Plapper (2019) sostiene que la preparación de datos es un paso clave para obtener la ventaja competitiva en la toma de decisiones previo a un razonamiento, puesto que, cumple con la finalidad de afrontar registros atípicos, faltantes, incoherentes e incompletos, resaltando así la importancia de la implementación de este en los flujos de trabajo de BA.

La limpieza de datos fundamentalmente se da porque en el paso de adquisición hay una alta probabilidad de tener datos duplicados, errados y faltantes. Dado el caso, se emplean un conjunto de pasos para la limpieza de estos, de acuerdo con la funcionalidad de los tipos de datos que se están trabajando.

Tableau (2020) recomienda algunos pasos para tener en cuenta en el procedimiento de limpieza. Asimismo, resalta que no existe un orden en general, ya que los datos presentan comportamientos distintos según los rubros en los que se encuentran. Estos pasos son: eliminación de datos duplicados, corrección de errores estructurales, filtrar y trata datos atípicos, y manejar datos faltantes.

Finalmente, el ordenamiento tiene la finalidad de estructurar los datos para poder facilitar su uso al momento de realizar análisis o creaciones de visualizaciones posteriores, logrando así una optimización en tiempo y recursos. Usualmente, el ordenamiento se basa en la separación por columnas y/o filas de las variables, agregación de columnas, cambio de filas por columnas, agrupaciones, creación de tablas anchas y/o largas, etc.

C. Razonamiento

Dentro del proceso de BA organizacional el razonamiento se divide en análisis descriptivo, predictivo y prescriptivo. Este análisis brindará un resultado previo a la toma de decisiones para generar un valor en la organización.

Análisis descriptivo

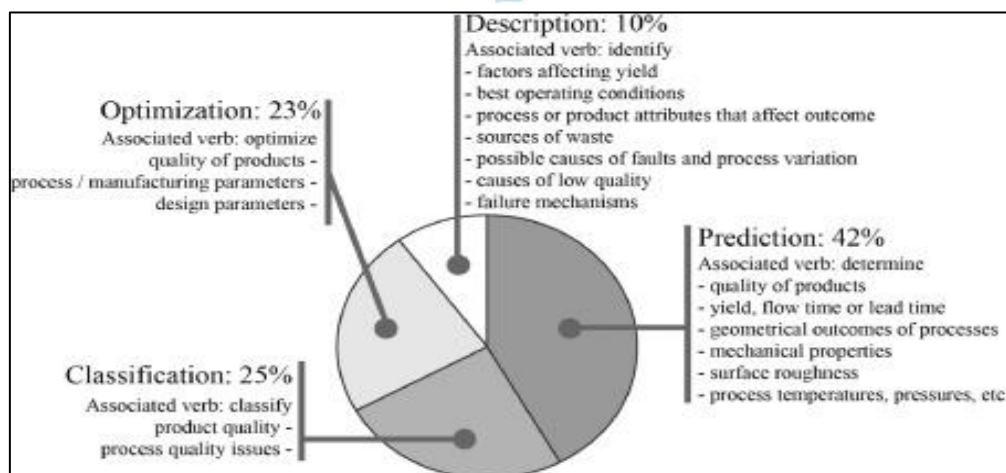
En general el análisis descriptivo se enfoca en dar una interpretación de datos pasados con la finalidad de entender, explicar y concluir el comportamiento de los parámetros analizados. Usualmente, es una práctica común entre las organizaciones. Omar, Minoufekar, & Plapper (2019) brinda un ejemplo claro de la aplicación de BA en el aseguramiento de calidad dentro de la gestión de una cadena de suministro, señalando que el análisis descriptivo abarca

un 10% de las tareas en ese proceso. Esta ejemplificación da el fundamento a su importancia porque este análisis implica conocer una serie de atribuciones que se encuentran perennes dentro de los procesos, como las mejores condiciones de trabajo, fuentes de residuos, fallas mecánicas, posibles causas de variaciones anómalas, etc. En síntesis, el objetivo principal del análisis descriptivo es dar una respuesta al cómo han ido funcionando los diversos indicadores o patrones en un determinado tiempo pasado.

Análisis predictivo

Este análisis tiene un enfoque en realizar predicciones en base al comportamiento estudiado en el análisis descriptivo. La variedad de aplicaciones dadas en los distintos rubros de las organizaciones responde a la importancia de implementar este análisis para dar respaldo a la toma de decisiones. Continuando con el ejemplo de la Figura 2 acerca de la calidad en la gestión de suministro; Omar, Minoufekar, & Plapper (2019) señala al análisis predictivo como el más resaltante, cubriendo un 42% de las tareas realizadas para optar por el aseguramiento de calidad, ejemplificando algunas se encuentran la determinación de las temperaturas y presiones para los procesos, calidad de productos, rendimientos y algunas otras propiedades mecánicas compatibles en el proceso. Fuera de este rubro, existen otras aplicaciones desarrolladas; Hindle, Kunc, Mortensen, Oztekin, & Vidgen (2019) describe algunos. Un claro ejemplo es la predicción de programación de citas en los hospitales, los cuales marcan un reto al momento de designar recursos para atenderlas porque varían según los requerimientos de los pacientes. Para ello, primero se enfocan en realizar todo un análisis descriptivo en los historiales de datos de los pacientes, con el fin de obtener la información necesaria para realizar un modelamiento predictivo y, así, obtener aproximaciones reales de próximos pacientes. De esta manera, es posible reducir tiempos de espera para la programación de citas.

Figura 2. Analítica en procesos de calidad.



Nota. Tomado de "Business analytics in manufacturing: Current trends, challenges and pathway to market leadership" Omar, Minoufekar, & Plapper (2019) .

Análisis prescriptivo

El análisis prescriptivo también tiene la orientación a generar una ventaja competitiva. Sin embargo, según Hindle, Kunc, Mortensen, Oztekin, & Vidgen (2019) este análisis cumple con un papel de generar opciones estratégicas al momento de tomar decisiones dentro de las organizaciones, marcando una diferencia con el análisis predictivo, el cual solo brinda información de un comportamiento futuro y, de acuerdo con ello, se puedan tomar decisiones. Es preciso indicar que, el análisis predictivo está integrado dentro de un prescriptivo.

D. Distribución.

Este último procedimiento consiste en brindar el producto de información al cliente final, siendo éste el conjunto de datos procesados o analizados según los requerimientos solicitados inicialmente. Nagle & Sammon (2016) hace referencia sobre la variabilidad de formatos y/o visualizaciones existentes al momento de distribuir la información final. Todos estos deben cumplir obligatoriamente con los requerimientos de los consumidores finales porque no tendría algún sentido dar otra información no solicitada.

Sun & Yen (2005) da a conocer un principio al momento de ofrecer una información final al usuario: los requisitos técnicos para utilizar la información deben ser menores a los que posee el usuario final. Adicionalmente, se debe tener en cuenta que la alfabetización digital del usuario debe tener la capacidad de entender la información brindada.

1.1.3 Cadena de suministro de información

Una cadena de suministro hace referencia a una serie de medios que empiezan desde la obtención de la materia prima hasta llegar al cliente según lo que desea. En tal sentido, Sun & Yen (2005) explica la cadena de suministro de información como una serie de agentes se encarga de reunir, interpretar y cumplir con los requerimientos de información del usuario, otorgándole esta al destinatario de la manera y en el momento correcto.

En el mismo contexto, Sun & Yen (2005) da a conocer el concepto de intercambio de información referenciándola como un conjunto de actividades entre partícipes de una organización con la finalidad de distribuir información. Además, plantea cuatro preguntas en las que debe estar basado este intercambio: qué compartir, a quién compartir, cómo compartir y cuándo compartir. En efecto, señala la importancia de responder correctamente estas preguntas para evitar deficiencias en el proceso de intercambio. Por tal motivo, enfatiza la implementación de un marco de *Information Supply chain* (ISC) con la finalidad de abordar y lograr resultados correctos en el intercambio de información. También, describe un enfoque interesante en los principios empleados en *Supply Chain Management* como lo es la implementación de los *Material Requirements Planning* (MRP), los cuales se trabajan en colaboración con los partícipes de la misma cadena y tienen el objetivo de evitar problemas

de suministro. Dentro de Information Supply Chain aplica la misma lógica, pero en este caso lo denomina *Informartion Requirement Planning* (IRP) el cual cumple con el mismo propósito de evitar deficiencias como duplicar, encontrar y enviar información repitente con la colaboración de los involucrados en la cadena de información.

Para complementar, el concepto y la importancia de ISC, Sun & Yen (2005) ejemplifica un caso dando a conocer el valor que tiene una correcta implementación de ISC junto a los métodos ya dados en SCM como lo es el Just in Time (JIT), cuyo propósito es tener un alto nivel de producción, pero manteniendo el mínimo inventario. Enfocándolo a ISC significa compartir la información correcta sin sobrecargarse. Por ejemplo, cuando una empresa de construcción solicita información del parámetro de lluvia del año 2017 de forma horaria a la Estación Científica Ramón Mugica (EC). La información correcta será brindar el total acumulado de lluvia por hora dentro del periodo y tiempo solicitado. Caso contrario sería, si el total acumulado de lluvia sería de otro año, de forma mensual y enviada luego de un prolongado tiempo, siendo esta totalmente obsoleta y sin cumplir los requerimientos del cliente.

1.2 Fenómeno el Niño y estaciones meteorológicas en Piura

El fenómeno “El Niño” (FEN) hace referencia a un evento climatológico acompañado de grandes precipitaciones siendo el causal de grandes pérdidas económicas y humanas. Por tal razón, a lo largo del tiempo se han ido implementando tecnologías para monitorear alteraciones climáticas que den indicios a la ocurrencia de este fenómeno como lo son la instalación de redes estaciones meteorológicas.

1.2.1 Fenómeno El Niño

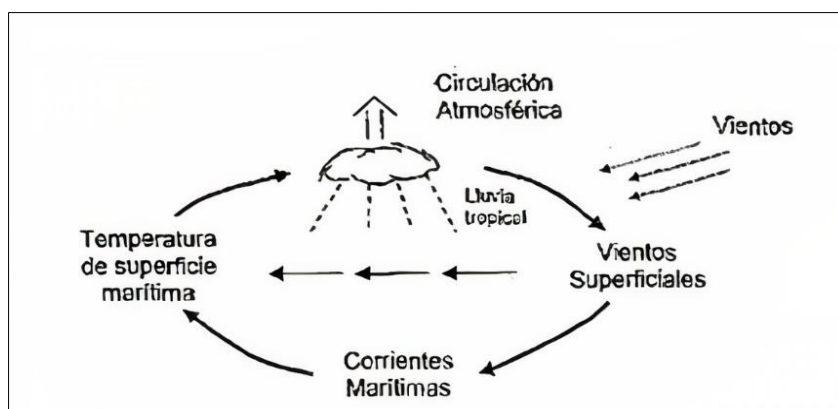
El FEN tiene varias definiciones, pero todas presentan términos en común. Takahashi (2014) afirma que las definiciones del FEN no se manejan de manera precisa en el Perú y el resto del mundo debido a la variedad de sus consecuencias. Él lo asocia a inundaciones en la costa norte del Perú a causa de un incremento de temperaturas en el océano Pacífico central.

Kovats, Bouma, Hajat, Worall, & Haines (2003) lo mencionan como un fenómeno oceánico atmosférico causado por las elevaciones de temperatura en la superficie del océano, debido al ingreso de corrientes cálidas en el pacífico ecuatorial desde la parte occidental, ocasionado por alteraciones en la presión atmosférica.

El Servicio Nacional de Meteorología e Hidrología del Perú (2014) lo cataloga como un modulador de variabilidad climática en todo el mundo. Adicionalmente, el Instituto del Mar del Perú (2018) señala que genera cambios en el ecosistema de la Corriente de Humboldt y en otras regiones del mundo debido a todas las teleconexiones. Además, influye en de forma directa en la circulación marina y procesos biogeoquímicos.

Zambrano (1986) lo resalta como un evento climático manifestado con alteraciones en las precipitaciones, temperatura del océano y aire, ocasionando variaciones en los ecosistemas marítimos, en los vientos y corrientes marinas cuyo mecanismo de interacción océano - atmosférico se ilustra en la Tabla 3.

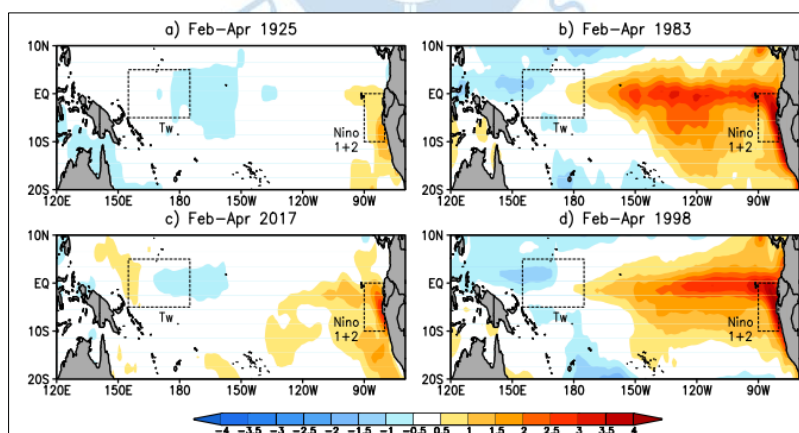
Figura 3. Mecanismo de interacción Océano Atmosfera en el Océano Pacifico Ecuatorial



Nota. Tomada de "El fenómeno de "El Niño" y la Oscilación Sur (ENSO)" Zambrano (1986).

La Figura 4, hace alusión a los principales eventos FEN en los últimos 100 años. En donde se aprecia el incremento de la temperatura superficial del mar en la parte del océano Pacífico central en relación al desarrollo de los FEN costeros y globales. Superando anomalías de más de 1.5°C respecto a las temperaturas de la superficie del mar promedio.

Figura 4. Variaciones de temperatura durante el Niño



Nota. Tomado de Fenómeno El Niño: "Global" vs "Costero" Takahash (2017).

Dentro de este marco de evento climático; Hajar, Bonilla, Munayco, Gutierrez, & Ramos (2016) recalcan la importancia de seguir desarrollando investigaciones de monitoreo y prevención para este fenómeno porque implica una tarea fundamental en la seguridad de salud pública porque garantiza la comprensión de medidas de prevención y vulnerabilidades a nivel local, regional y nacional.

1.2.2 Cadena de estaciones meteorológicas Niño 5

En la formación del FEN se hace seguimiento de una serie de parámetros, categorizándose en dos tipos (oceánicos y atmosféricos) los cuales tienen la característica de tener perceptibles anomalías previas a un evento climatológico. La Tabla 1 muestra estos parámetros.

Tabla 1. Parámetros océano – atmosféricos del FEN

Oceánicos	Unidades	Atmosféricos	Unidades
Temperatura superficial	(°C)	Presiones atmosféricas	(mbar)
Temperatura sub-superficial	(°C)	Vientos de superficie	(m/s y azimuth)
Profundidad de termoclina	(m)	Vientos de altura	(m/s y azimuth)
Nivel de la superficie del mar	(cm)	Humedad	%
Corrientes marinas	(m/s)	Nubosidad	%
Ondas	(m/s)	Precipitaciones	mm

Nota. Adaptado de “El Fenómeno El Niño- Oscilación sur” Rodríguez (2011).

Aunado a lo mencionado anteriormente, en el año 2009 se aprobó un proyecto presentado por el Laboratorio de Física de la Universidad de Piura, denominado “Niño5”, siendo una cadena de estaciones meteorológica para realizar investigaciones de los parámetros atmosféricos entre la costa y sierra norte del Perú. Los aliados en este proyecto fueron el FINCYTC, Gobierno Regional de Piura y la Universidad de Piura.

Las estaciones meteorológicas de la cadena NIÑO5 se encuentran instaladas en lugares cercanos a 5 grados de latitud sur. Actualmente está conformada por 10 estaciones cuyas ubicaciones y estados se indican en la Tabla 2. Las estaciones de Bayovar, Cerritos y Poclús son las últimas instaladas.

Tabla 2. Ubicación de las estaciones meteorológicas que conforman la cadena NIÑO5

N°	Lugar	Distrito	Provincia	Cuenca	Latitud(S)	Longitud(O)	Altitud	Estado
1	Balcones	La Brea	Talara	Pacífico	04°40′ 11.70″	81° 19′ 38.00″	18	No operativo
2	Colán	Paíta	Paíta	Pacífico	05°03′ 15.30″	81° 03′ 43.20″	79	Operativo
3	Piura	Piura	Piura	Piura	05°10′ 14.00″	80° 38′ 18.60″	45	Operativo
4	Chávez	Chulucanas	Chulucanas	Piura	05°11′ 22.58″	79° 58′ 01.16″	175	Operativo
5	Morropón	Morropón	Morropón	Piura	05°11′ 22.58″	79° 58′ 01.16″	131	No operativo
6	Huancabamba	Huancab.	Huancab.	Huancab.	05°14′ 12.00″	79°26′ 54.00″	1957	Operativo
7	Bayovar	Bayovar	Sechura	Pacífico	05°47′ 39.00″	81°04′ 14.20″	75	Operativo
8	San Ignacio	San Ignacio	San Ignacio	Huancab.	05°09′ 01.00″	78°59′ 45.90	1270	Operativo
9	Cerritos	Cristo Nos Valga	Sechura	Piura	05°29′ 55.64″	80°43′ 08.65″	16	Operativo
10	Alto de Poclús	Frias	Ayabaca	Piura	04°55′ 47.00″	79°53′ 34.00″		Operativo

La Figura 5 muestra la distribución de las estaciones meteorológicas en un mapa geográfico.

Figura 5. Mapa de ubicaciones de estaciones meteorológicas de la cadena NIÑO5.



Rodríguez (2013) señala que el objetivo de la cadena NIÑO5 es generar una base de datos meteorológicos con el fin de ayudar en investigaciones regionales en las áreas de meteorología, oceanografía, producción agrícola y pesquera, energías alternativas, además de realizar predicciones y prevenciones dentro de estas áreas.

La Tabla 3 muestra las aplicaciones de estas estaciones meteorológicas, seccionándolas de 6 maneras: Climáticas, oceánicas, ecológicas, productivas, energéticas y preventivas.

Tabla 3. Aplicaciones de estudio con datos registrados por la cadena NIÑOS

Tipo	Aplicación
Climáticas	Cuantificación de alteraciones ambientales debido a la ocurrencia del FEN
	Diferencia de los efectos del FEN en la parte costa y sierra
	Efectos de cambio climático en costa y sierra.
	Mecanismo del fenómeno trasvase y su incidencia en la ocurrencia de lluvias
Oceánicas	Estudio del fenómeno del aloramiento costero y su impacto en el ecosistema marino-costero.
Ecológicas	Variaciones ecológicas Costa - Sierra
	Seguimiento ecológico y climático del ecosistema en la región Piura
Productivas	Incidencia de las condiciones ambientales en la producción agrícola
	Incidencia de las condiciones marino-costeras en la producción pesquera
Energéticas	Potencial regional de energía eólica
	Potencial regional de energía solar
Preventivas	Incidencia de las condiciones ambientales en la Salud de la población
	Prevención y aleta de avenidas del río Piura.

Nota. Adaptado de “Seis estaciones meteorológicas para el estudio de los efectos del fenómeno climático El Niño en las actividades productivas y en la población” Rodríguez & Marcelo (2009).

Estas estaciones registran parámetros en común, como a su vez tienen distintos sensores cuya función es obtener datos para cada parámetro en específico. (Ver Tabla 4)

Tabla 4. Parámetros medidos en la cadena NIÑO5

Parámetro	Medidor/sensor	Unidad
Temperatura	Termómetro	°C
Presión atmosférica	Barómetro	mB
Humedad Relativa	Higrómetro	%
Velocidad del viento	Anemómetro	m/s
Dirección del viento		Ángulo Azimuth
Radiación solar	Piranómetro	W/m ²
Lluvia	Pluviómetro	mm

Nota. Adaptado de “Seis estaciones meteorológicas para el estudio de los efectos del fenómeno climático El Niño en las actividades productivas y en la población” Rodríguez & Marcelo (2009).

1.3 Cadena de suministro de información en la cadena NIÑO5

El fin de la cadena NIÑO5 es proporcionar datos de los parámetros registrados en cada una. Para realizar un análisis de los datos se atraviesa por una serie de procedimientos desde la obtención hasta su preparación, en cada uno se encuentran presentes distintos retos que originan una mayor dedicación de tiempo a la preparación que al análisis final.

1.3.1 Estado actual

La cadena NIÑO5 registra valores de los parámetros indicados en la Tabla 4. Es preciso mencionar que cada una de las estaciones tiene configuraciones diferentes porque son marcas de fabricante distintas entre sí y poseen configuraciones desiguales.

La Tabla 5 muestra el número de parámetros registrados para cada una de las estaciones. Todas a excepción de la estación Chávez registran más de 30 parámetros. Dentro de estos, se encuentran parámetros secundarios los cuales dependen de los valores de otros registros o simplemente son configuraciones agregadas por la misma marca de los registradores.

La Tabla 5 se ha subdividido en 6 parámetros: temperatura, humedad, presión, lluvia acumulada, viento y radiación solar. Todas las estaciones registran estos parámetros. Con respecto a la temperatura, hay estaciones como la de Colán y Chávez que también miden la temperatura del suelo. Otras poseen variables secundarias. Por ejemplo, máximos, mínimos o algún otro cálculo derivado de la temperatura. En humedad, las estaciones de Colán, UDEP y Chávez también registran el punto de rocío, mientras que las estaciones Huancabamba y San Ignacio solo registran variables secundarias extras a la humedad. Para los parámetros de presión atmosférica y lluvia acumulada, la estación Chávez no registra esta primera. De igual manera, tampoco registra dirección del viento. Finalmente, los parámetros asociados a energía solar lo poseen todas las estaciones. Sin embargo, las estaciones Colán y UDEP sobresalen en parámetros porque también registra niveles de indicadores UV.



Tabla 5. Parámetros registrados por cada estación meteorológica

Nombre	N°	Parámetros												Obtención	
		°C	Descrip.	HR	Descrip.	Bar	Descrip.	Rain	Descrip.	Viento	Descrip.	Solar	Descrip.		Secund.
Colán	38	6	°C aire	2	HR.	1	mBar	1	Rain	3	Velocidad	14	Radiación	11	Manual
			°C suelo		Pto. rocío						Orientación		Índices UV		
UDEP	31	3	°C aire	2	HR	1	mBar	2	Rain	6	Velocidad	6	Radiación	11	Manual
			Max, Min		Pto. rocío				Prom.		Orientación		Índices UV		
Chávez	8	2	°C aire	2	HR	-		1	Rain	2	Velocidad	1	Radiación	-	Online
			°C suelo		Pto. rocío						Orientación				
Huancab.	52	4	°C aire	4	HR	4	mBar	4	Rain	8	Velocidad, orientación	4	Radiación	24	Manual
			Max, min		Max, min		Max, min		Max, min		Max, min		Max, min		
San Ignacio	52	4	°C aire	4	HR	4	mBar	4		8	Velocidad, orientación	4	Max y Min	24	Manual
			Max, min		Max, min		Max, min		Max, min		Max y min				

1.3.2 Problemática

Los datos de la Cadena NIÑO5 son requeridos de manera frecuente por instituciones y empresas interesadas, quienes solicitan datos en rango de tiempo mayores a 1 año y en frecuencias horarias, diarias y/o mensuales, con fines de investigación para el desarrollo de diversos proyectos, ya sean de infraestructura, salud, trabajos de grado, informes como anuarios, entre otros. Los datos también son suministros para investigaciones que se realizan dentro de la UDEP.

La problemática se evidencia al momento de realizar la preparación de las fuentes de datos. El procedimiento para la preparación se realiza manualmente, utilizando hojas de cálculo Excel y, en otros casos, a través de algoritmos implementados en los lenguajes de R o Python. Esto origina un cuello de botella en este procedimiento porque se demanda un tiempo mayor, el cual podría destinarse al procedimiento analítico y no en gran porcentaje a la preparación.

La Figura 6 muestra un caso específico relacionado a lo mencionado anteriormente. En esta ocasión, la tabla hecha por una tesista muestra datos recolectados de radiación solar de la estación de Colán para hacer un análisis del potencial solar en esta localidad. Estos se han registrado de forma manual en una hoja de cálculo Excel. Las principales desventajas de aplicar estos procedimientos es la verificación de la existencia de datos errados, faltantes o duplicados. En el caso que se requieran recopilar datos de otros parámetros y/o estaciones en conjunto, la tarea se dificulta en un nivel más alto, retrasando el procedimiento del análisis. El tiempo promedio empleado para este procedimiento es de 15 a 20 horas para trabajar con más de 10 años de datos.

Figura 6. Matriz de datos radiación solar de la estación Colán

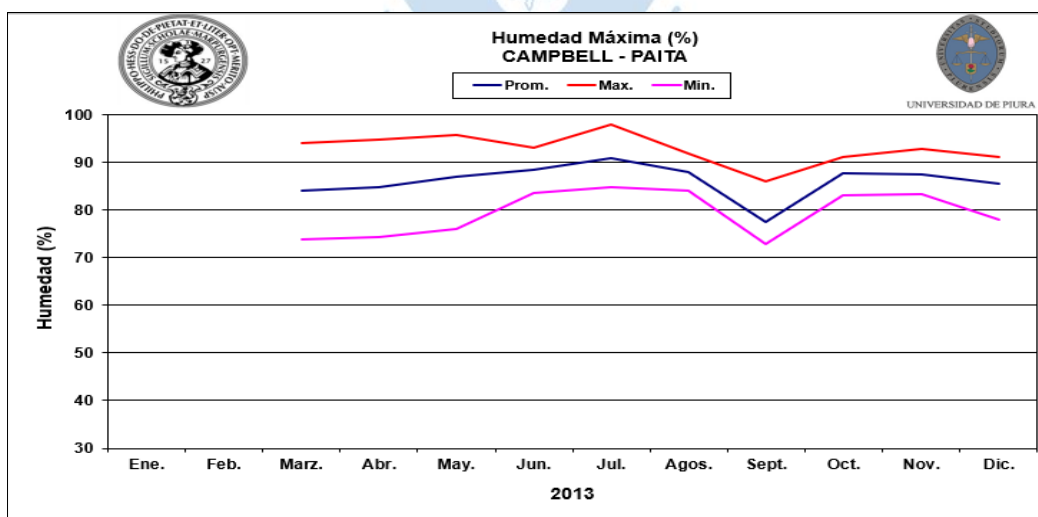
A	B	C	D	E	F	G	H	I	J	K	L
Item	DIA/MES-HORA	RADIACION (W/m2)									
		2013	2014	2015	2016	2017	2018	2019	2020	2021	PromRS
31	01/01 05:00		-0.599	-0.247	-1.108	-0.922	-2.013				-0.978
32	01/01 05:10		-0.596	-0.275	-1.047	-0.871	-1.861				-0.930
33	01/01 05:20		-0.672	-0.249	-0.75	-0.589	-1.663				-0.785
34	01/01 05:30		-0.845	-0.257	-0.622	-0.657	-0.829				-0.642
35	01/01 05:40		-0.809	-0.217	-0.517	-0.712	-0.427				-0.536
36	01/01 05:50		-1.037	-0.271	-0.406	-0.78	-0.43				-0.585
37	01/01 06:00		-1.644	-0.242	-0.337	-0.828	-0.649				-0.740
38	01/01 06:10		-0.593	0.224	0.118	-0.569	-1				-0.364
39	01/01 06:20		1.863	1.684	2.031	0.197	-0.033				1.148
40	01/01 06:30		5.396	5.446	6.729	2.034	12.54				6.429
41	01/01 06:40		16.31	11.96	14.33	9.34	35.69				17.526
42	01/01 06:50		32.71	19.52	23.05	19.23	69.59				32.820
43	01/01 07:00		45.78	26.37	33.09	30.4	118.4				50.808
44	01/01 07:10		59.45	40.67	43.66	50.56	152.4				69.348
45	01/01 07:20		74.61	45.62	59.68	66.65	194.1				88.132
46	01/01 07:30		99.5	45.52	82	82.2	251.2				112.084
47	01/01 07:40		115.1	57.54	99.8	110.1	279.8				132.468
48	01/01 07:50		144.6	72.11	121	112.2	273				144.582
49	01/01 08:00		178.7	102.4	144.9	125.4	296.7				169.620

La Figura 7 muestra una tabla de datos realizados en hojas Excel para la publicación de anuarios en el año 2013 para la estación de Colán, específicamente en el parámetro de humedad relativa en una frecuencia diaria. Este procedimiento se repite para todos los parámetros. De la misma forma, la Figura 8 muestra la creación de las visualizaciones en el mismo libro de trabajo. Adicionalmente, la elaboración de anuarios utilizando Excel toma un tiempo de 20 a 25 horas por estación según el director de la EC.

Figura 7. Ejemplo de datos de humedad relativa de la estación de Colan, año 2013.

B	C	D	E	F	G	H	I	J	K	L	M	N
ESTACION METEOROLOGICA CAMPBELL - PAITA												
PARAMETRO: HUMEDAD RELATIVA MÁXIMA												
UNIDAD: %												
AÑO: 2013												
Dia	Ene.	Feb.	Marz.	Abr.	Mag.	Jun.	Jul.	Agos.	Sept.	Oct.	Nov.	Dic.
1				89.00	87.20	86.20	94.10	86.30	82.10	88.50	88.00	88.50
2				81.10	83.60	86.90	91.50	87.30	77.24	89.00	86.30	90.30
3				83.40	87.30	89.10	94.50	87.00	83.21	88.60	88.70	90.60
4				75.56	87.10	85.00	91.60	90.30	78.19	83.00	87.80	79.40
5				74.42	84.10	86.50	92.00	86.00	79.08	86.10	87.90	83.10
6				85.20	93.00	93.10	91.60	88.70	85.94	85.00	87.80	87.20
7				85.30	84.40	92.80	87.80	89.70	81.32	89.80	87.70	87.30
8				81.20	88.50	92.80	90.50	88.50	77.32	86.70	89.90	86.60
9			80.30	75.12	84.50	85.30	94.50	90.20	79.03	84.60	91.40	86.30
10			84.80	83.70	84.10	84.00	89.20	87.90	78.06	87.00	84.20	83.10
11			73.72	74.23	86.10	85.40	85.80	85.00	81.08	90.50	83.70	87.70
12			74.25	80.40	84.80	87.70	87.80	87.40	77.14	88.00	91.70	84.90
13			82.90	83.70	82.20	85.90	91.00	85.90	77.67	90.90	83.70	88.00
14			91.90	85.20	89.90	88.70	88.60	86.90	76.14	89.10	90.70	79.33
15			79.54	84.60	93.70	87.40	90.50	87.30	72.80	88.50	88.80	83.10
16			80.80	81.10	90.10	89.70	91.10	89.60	73.06	91.20	86.10	84.30
17			84.50	86.50	76.03	92.30	92.80	89.50	77.42	86.50	83.40	87.80
18			88.50	90.10	86.20	87.60	95.00	91.90	76.54	88.60	84.80	90.30
19			79.38	83.00	84.20	90.10	96.90	91.90	74.66	87.00	85.80	87.30
20			79.75	84.50	92.90	86.50	92.00	88.90	74.67	85.70	86.20	91.20
21			87.50	91.40	95.80	86.40	84.80	85.50	75.67	88.00	88.10	87.80
22			89.40	87.30	92.00	90.60	90.80	84.90	73.62	86.60	89.30	87.10
23			82.90	85.20	88.30	87.70	88.40	84.00	76.58	88.60	84.90	89.20
24			84.90	85.60	82.70	91.40	87.00	88.90	80.66	87.50	87.30	79.98
25			86.70	88.90	84.90	89.80	87.50	88.30	76.29	89.90	87.90	88.50
26			86.20	89.30	84.90	89.10	88.70	86.90	74.26	89.00	92.90	82.60
27			94.10	88.40	84.10	90.10	85.80	85.70	74.07	87.80	85.00	78.09
28			84.30	84.60	83.70	89.20	96.60	88.50	80.01	86.00	86.00	81.20

Figura 8. Gráfico de variación mensual de la humedad relativa de la estación Colán, año 2013



Por otro lado, la Figura 9 muestra una matriz con un análisis general basado en porcentajes de datos nulos y atípicos para la estación de Colán. Este mismo comportamiento se encuentra presente en las otras estaciones. Demostrando así el desafío en el procedimiento de preparación de datos al momento de identificarlos. Sin embargo, la naturaleza de estos datos siempre estará presente por causas fuera de control como fallas técnicas en los *Data logger* o en los sensores afectados tal vez por un evento climático.

Figura 9. Matriz de datos anómalos y nulos (null) para la estación Colán.

Medida	Tipo de anomalía	Observación	Cantidad	Inicio	Fin	%
Temperatura	Null		13654	1/01/2013	9/03/2013	3.34%
				18/05/2014	30/05/2014	
				1/06/2014	13/06/2014	
				25/08/2014	1/09/2014	
Humedad relativa	Null		13578	1/01/2013	9/03/2013	3.32%
				18/05/2014	30/05/2014	
				1/06/2014	13/06/2014	
				28/08/2014	30/08/2014	
WindDir	Null		15001	1/01/2013	9/03/2013	3.67%
				19/05/2014	9/06/2014	
				28/08/2014	30/08/2014	
RadiacionSolar	Null		13578	1/01/2013	9/03/2013	3.32%
				18/05/2014	30/05/2014	
				1/06/2014	13/06/2014	
				28/08/2014	30/08/2014	
	Anomalía	7999 w/m2	3334	4/08/2014	25/08/2014	0.82%
Rain	Null		13584	1/01/2013	9/03/2013	3.32%
				19/05/2014	06/05/20104	
WindSpeed	Anomalía	(100-1100) m/s	10430	1/01/2015	1/01/2016	2.55%
				1/01/2016	1/01/2017	
				1/01/2017	1/01/2018	
				1/01/2018	1/01/2019	
				1/01/2019	1/01/2020	
				1/01/2013	9/03/2013	
	Null		15004	19/05/2014	16/06/2014	3.67%
				28/08/2014	30/08/2014	
				1/01/2015	1/01/2016	

1.3.3 Propuesta de solución

Teniendo en cuenta la deficiencia originada en el procedimiento de preparación, se propone implementar una cadena de suministro de información. Con la finalidad de proporcionar datos y visualizaciones según los requerimientos finales de los usuarios para que ellos puedan acceder a ellos de una manera online, siempre y cuando se les permita el acceso solicitado.

Dentro de la cadena de suministro de información habrá agentes, trabajadores de la estación científica Ramón Mugica de la Universidad de Piura, quienes desarrollarán la preparación mediante pasos que se definirán en este trabajo, los cuales estarán basados en los principios de integración, limpieza y ordenamiento.

La implementación de esta solución va acompañada del trabajo en el software Tableau, el cual permitirá la preparación, visualización y distribución de los datos a los usuarios interesados e involucrados.



Capítulo 2

Metodología

2.1 Obtención y descripción de fuente de datos.

Las estaciones de la cadena NIÑO5 con las que se trabajarán son las de Colán, UDEP, Chávez, Huancabamba y San Ignacio. Se realizará la descripción correspondiente basadas en su frecuencia de registro, rango de años, nombres de parámetros y tipos de archivos.

Se trabajará con los parámetros en común de las estaciones meteorológicas mencionadas anteriormente. Estos son: temperatura del aire, humedad relativa, velocidad del viento, dirección del viento, lluvia acumulada, radiación solar y presión atmosférica. La Tabla 6 muestra el nombre con que se encuentran registrados estos parámetros para cada estación, además de las unidades de medida para cada una de estas. Debido a que, el parámetro de orientación de viento de la estación UDEP difiere en unidades con otras, se optará por el cambio a unidades sexagesimales según los valores de la Tabla 7.

Tabla 6. Identificación de parámetros

ID	Estación Meteoreológica	Temperatura	Unidad	Humedad Relativa	Unidad	Lluvia acumulada	Unidad	Velocidad del viento	Unidad	Dirección del viento	Unidad	Radiación Solar	Unidad	Presión	Unidad
1	Colán	AirTC_Avg	°C	RH	%	Rain_m m_tot	mm	WS_ms _Avg	m/s	Wind Dir	Grados sexagesimales	Sr100U p_Avg	w/m ²	Bp_mb ar_Avg	mBar
2	UDEP	Hi Temp	°C	Out Hum	%	rain	mm	Out Speed	m/s	Hi Dir	Abreviatura de Rosa de vientos	Hi Solar Rad.	w/m ²	Bar	mBar
3	Chavez	Temp, °C	°C	HR, %	%	Lluvia, mm	mm	Velocidad del viento	m/s	-		Radiación Solar	w/m ²	-	mBar
4	Huancabamba	CH9 AVG	°C	CH12 AVG	%	CH4 AVG	mm	CH1 AVG	m/s	CH7 AVG	Grados sexagesimales	CH10 AVG	w/m ²	CH11A VG	mBar
5	San Ignacio	CH9 AVG	°C	CH12 AVG	%	CH4 AVG	mm	CH1 AVG	m/s	CH7 AVG	Grados sexagesimales	CH10 AVG	w/m ²	CH11A VG	mBar

Tabla 7. Valores sexagesimales en dirección del viento.

Orientación	Valor	Rango
N	359°	<348.75 – 359] y <0 – 11.25]
NNE	22.5°	<11.25 -33.75]
NE	45°	<33.75 - 56.25]
ENE	67.5°	<56.25 - 78.75]
E	90°	<78.75 – 101.25]
ESE	112.5°	<101.25 – 123.75]
SE	135°	<123.75 – 146.25]
SSE	157.5°	<146.25 – 168.75]
S	180°	<168.75 – 191.25]
SSW	202.5°	<191.25 – 213.75]
SW	225°	<213.75 – 236.25]
WSW	247.5	<236.25 – 258.75]
W	270°	<258.75 – 281.25]
WN	292.5°	<281.25 – 303.75]
NW	315°	<303.75 – 326.25]
NNW	337.5	<326.75 – 348.75]

La Tabla 8 muestra el resumen de las fuentes de datos obtenidas correspondiente a cada estación meteorológica. La columna fechas de registro hace referencia al rango de fechas de datos obtenidos de inicio a fin. Mientras que la columna rangos de fecha y frecuencia de registro indica el intervalo de fechas para cada frecuencia de registro en minutos. Por ejemplo, la estación UDEP (con ID 2) el rango de fechas obtenido es desde el 7 de mayo del 2010 desde las 8 horas con 30 minutos hasta el 1 de enero del 2022 a las 00 horas. Luego, desde el 7 de mayo del 2010 hasta el 28 de febrero del 2017 se registraban datos cada 30 minutos, luego a partir de esta última fecha al 1 de enero del 2022 se registran cada 10 minutos.

Tabla 8. Resumen de las fuentes de datos con sus respectivos registros

ID	Registro		Rangos de fechas y frecuencia de lectura (min)		
	Inicio	Fin	60	30	10
1	01/01/2013	01/07/2022	-	-	01/01/2013
	00:10	00:00			-
					01/07/2022
2	07/05/2010	01/01/2022	-	07/05/2010	28/02/2017
	08:30	00:00		-	-
				28/02/2017	01/01/2022
3	23/05/2010	15/10/2021	23/05/2010	24/05/2016	05/06/2017
	10:00	15:00	-	-	-
			24/05/2016	05/06/2017	15/10/2021
4	27/07/2011	30/07/2021	-	-	27/07/2011
	11:40	08:50			-
					30/07/2021
5	31/08/2012	31/07/2021	-	-	31/08/2012
	00:00	00:00			-
					31/07/2021

La Tabla 9 resumen los nombres en la que las fuentes de datos se encuentran registradas con su respectivo tipo de archivo. Como se observa, el tipo de archivos no es uniforme para cada uno, por lo tanto, todos estos se convertirán a formato CSV. Finalmente, la estructura de todos los archivos de datos deberá ser igual. Estos deben iniciar con el nombre de los parámetros en la primera fila.

Tabla 9. Formatos de los archivos de datos

ID	Archivos	
	Nombre	Tipo
1	Cr1000Colan(aaaa/mm/dd)	TXT y DAT
2	UDEP (aaaa/mm)	TXT
3	Chavez(aaaa/mm/dd)	TXT, CSV y XLSX
4	HBBA(aaaa)	TXT
5	SanIgnacio(aaaa)	TXT

2.2 Software

Se empleará el software de la empresa Tableau para la implementación de la cadena de suministro de información. Tableau provee el software para realizar BA. En este trabajo de investigación se utilizaron los productos: Tableau PrepBuilder, Tableau Desktop y Tableau Online.

2.2.1 Tableau Prep Builder

Tableau PrepBuilder es una herramienta diseñada exclusivamente para la preparación de datos de una manera interactiva e intuitiva con el usuario. Dentro de esta existen diferentes mecanismos para limpiar y ordenar los datos optimizando el tiempo en la preparación. Se caracteriza por facilitar la conexión a distintas fuentes de datos en diferentes formatos. Asimismo, permite al usuario tener un panorama visual inmediato respecto a los pasos que va empleando y las modificaciones que se van originando de forma paralela. Este también permite exportar los datos según las características solicitadas.

2.2.2 Tabpy

Tableau PrepBuilder cuenta con una extensión dentro de sus flujos para trabajar con Scripts de Python (Tabpy) permitiendo así su conexión con archivos locales de este tipo. La ventaja ofrecida es el uso de algoritmos para fines de limpieza o predicción, de acuerdo con los requerimientos de análisis que el usuario desea utilizar.

2.2.3 Tableau Desktop

Tableau Desktop permite al usuario conectarse a una fuente de datos con la finalidad de crear visualizaciones, según los comportamientos que se desean analizar para cubrir necesidades de razonamiento, sobre todo de análisis descriptivo. Con esta herramienta se puede realizar una exploración de los datos de una manera visual empleando mecanismos interactivos logrando así la explotación eficiente de los datos. Además, permite insertar una

variedad de cálculos visuales complementarios con el objetivo de obtener más información de manera analítica.

2.2.4 Tableau Online

Tableau Online es una herramienta insertada en la nube, en donde permite a los distintos miembros agregados publicar hojas de trabajo visuales y/o flujos de preparación de datos. Adicionalmente, permite trabajar con ellos mediante dispositivos conectados a internet. Asimismo, tiene la capacidad de poder conectarse a una variedad de fuente de datos en nube o actualizados por los miembros, según los permisos y limitaciones brindadas inicialmente.

2.3 Preparación de datos

En la preparación de datos se establecen los pasos y criterios para el desarrollo de las tareas de limpieza y ordenamiento de todas las fuentes de datos obtenidas de las estaciones meteorológicas.

2.3.1 Limpieza

En primer lugar, se realizará la tarea de integración de todos los datos en la herramienta de Tableau PrepBuilder. Cada estación debe tener un flujo de integración en donde se encuentren importados todos los archivos de las fuentes de datos correspondientes. Luego, se les asigna un nombre y formato a los parámetros determinados anteriormente, los cuales deberán ser utilizados para todos los flujos. (Ver Tabla 10)

Tabla 10. Nombre y formato de los valores de los parámetros

Parámetro	Nombre	Formato	Ejemplo
Temperatura del aire	Temp_air	Decimal	TT.T
Humedad Relativa	RH_%	Decimal	HH.H
Radiación Solar	Radiation_Solar	Decimal	RRRR.R
Velocidad de viento	Wind_Speed	Decimal	VV.V
Dirección del viento	Wind_Dir	Entero	WWW
Presión Atmosférica	Presión_Bar	Decimal	PPPP.P
Lluvia acumulada	Rain_mm	Decimal	MM,M
Fecha y hora	DATETIME	Fecha y hora	dd/mm/aaaa hh:mm

Cumplido con lo anterior, se establecerán los criterios para el tratamiento de los datos nulos y atípicos. Primero, para decidir si un dato es atípico, deberán cumplir con las características establecidas en la Tabla 11, las cuales han sido previamente verificados y aprobados por el director de la estación científica. Una vez identificados, estos se convertirán en datos nulos.

Finalmente, la limpieza de los datos nulos pasará por un procedimiento de reemplazos. Este se trabajará en un script de limpieza Python, utilizando la herramienta de extensión *Tabpy*. Los nulos se reemplazarán por los valores con el promedio de los registros anterior y posterior a la posición donde se encuentra el valor a reemplazar. Cabe precisar que este criterio puede cambiar según lo que requiera el usuario al momento de preparar la data. Recaltar que cada etapa terminará en un archivo exportado en formato CSV.

Tabla 11. Rango de valores de datos atípicos

Parámetros	Colán	Udep	Chávez	Huancabamba	San Ignacio
Temperatura (°C)	< 10	< 10	< 10	< 8	< 8
	>40	>40	>40	>40	>40
Humedad relativa (%)	>30	>30	>30	>30	>30
	<101	<101	<101	<101	<101
Radiación solar (w/m2)	< 0	< 0	< 0	< 0	< 0
	>1200	>1200	>1200	>1200	>1200
Velocidad del viento (m/s)	>20	>20	>20	>20	>20
Orientación del viento (°)	< 0	< 0	-	< 0	< 0
	>360	>360		>360	>360
Presión atmosférica (mBar)	<1000	<1000	-	< 790	< 865
	>1120	>1120		>810	>885
Lluvia acumulada(mm/h)	>50	>50	>50	>50	>50

2.3.2 Ordenamiento

Una vez finalizada la limpieza, de igual manera se trabajará con flujos de preparación para cada estación meteorológica importando el archivo final CSV de la etapa anterior. En primer lugar, se determinará la agrupación de los datos en forma horaria. Es decir, como existen registros cada 10,30 y 60 minutos, todos estos se agruparán por horas. Luego, se introducirán campos agregados para cada parámetro según sus criterios correspondientes, tal como lo muestra la Tabla 12. Cabe recalcar que el director de la estación científica ha revisado y aprobado estos criterios. Al finalizar este flujo, los datos deberán ser exportados a un archivo CSV.

Tabla 12. Criterios de agregación

Parámetro	Criterio de agregación	
Temperatura del aire	Máxima, Mínima	Promedio,
Humedad Relativa	Máxima, Mínima	Promedio,
Radiación Solar	Máxima, Promedio	
Velocidad de viento	Máxima, Promedio	
Orientación del viento	Promedio	
Presión Atmosférica	Máxima, Mínima	Promedio,
Lluvia acumulada	Sumatoria	

Finalmente, se trabajará con un flujo de integración en donde importará los archivos finales de la etapa anterior de cada una de las estaciones meteorológicas. Estos datos también se exportarán en formato CSV.

2.4 Análisis y distribución de información

Una vez terminada el procedimiento de preparación, se desarrollarán las visualizaciones con la herramienta de Tableau Desktop para presentar un análisis descriptivo visual. Estas visualizaciones deberán cumplir criterios de desarrollo aprobadas por el director de la estación científica. (Ver Tabla 13)

Posterior a estas visualizaciones individuales, todas se incorporarán a dos Dashboard donde deberán cumplir con los requisitos de interactividad, información entendible, rápida y eficaz. Este, deberá filtrarse a través de los años de registro y las estaciones que el usuario desea ver.

Finalmente, la distribución de todos los flujos de preparación y visualizaciones serán mediante la herramienta de Tableau Online. Las hojas de trabajo desarrolladas serán subidas a esta plataforma. Las fuentes de datos se actualizarán cada 6 meses, según los datos que se vayan recopilando. Dentro de los miembros con acceso a esta plataforma estarán los trabajadores y tesistas de la estación científica, quienes tendrán el permiso de poder editar y trabajar con estos flujos.

2.4.1 Diseño e implementación de visualizaciones

Una vez terminada el procedimiento de preparación, se desarrollarán las visualizaciones con la herramienta de Tableau Desktop para presentar un análisis descriptivo visual. Estas visualizaciones deberán cumplir criterios de desarrollo aprobadas por el director de la estación científica, como se muestran en la Tabla 13 para cada parámetro.

Tabla 13. Criterios de visualizaciones

Parámetro	Gráficos	Periodos de resumen
Temperatura del aire	Climograma	Mensual
Humedad Relativa	Líneas de Máxima, Promedio, Mínima	Mensual
Radiación Solar	Radiación solar horario y energía solar mensual	Horaria y mensual
Velocidad de viento	Líneas de Máxima promedio, promedio y frecuencia de vientos	Mensual
Orientación del viento	Frecuencias de orientaciones y velocidades	Total de años
Presión Atmosférica	Líneas de Máxima, Promedio, Mínima	Mensual
Lluvia acumulada	Climograma	Mensual

Posterior a estas visualizaciones individuales, todas se incorporarán a dos *Dashboard* donde deberán cumplir con los requisitos de interactividad, información entendible, rápida y eficaz. Este, deberá filtrarse a través de los años de registro y las estaciones que el usuario desea ver.

Finalmente, la distribución de todos los flujos de preparación y visualizaciones serán mediante la herramienta de Tableau Online. Las hojas de trabajo desarrolladas serán subidas a esta plataforma. Las fuentes de datos se actualizarán cada 6 meses, según los datos que se vayan recopilando. Dentro de los miembros con acceso a esta plataforma estarán los trabajadores y tesistas de la estación científica, quienes tendrán el permiso de poder editar y trabajar con estos flujos.

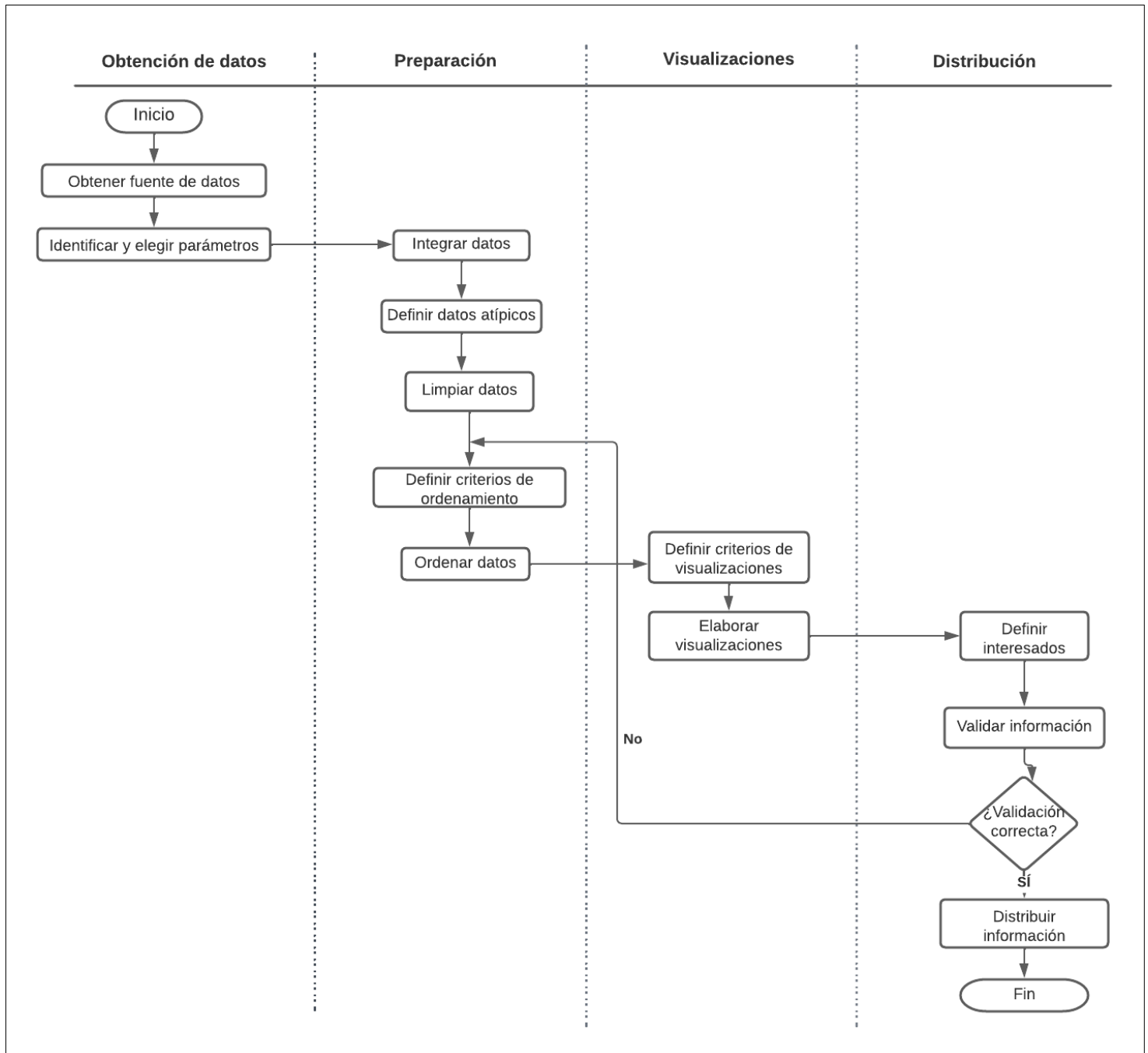
2.4.2 Validación

La etapa de validación se realiza a través de dos mecanismos. En primer lugar, los dashboard resueltos deberán ser aprobados por los asesores del presente trabajo, quienes darán las observaciones y la conformidad final. Posteriormente, se presentará el procedimiento de la implementación de la cadena de suministro de información, a personas externas a la estación científica, con el objetivo de verificar la comprensión de la información y la interactividad con ella. Mínimo se presentará a un grupo de visita conformado por estudiantes universitarios y especialistas en el tema de meteorología.

Con referencia a la distribución de la información en Tableau online, el director de estación científica dará acceso con permisos de visualización y descarga de datos a las personas que considere como interesados y miembros.

La Figura 10 muestra de manera general los pasos a realizar para la correcta implementación de la cadena de suministro de información en la red de estaciones

Figura 10. Diagrama de flujo cadena de suministro de información



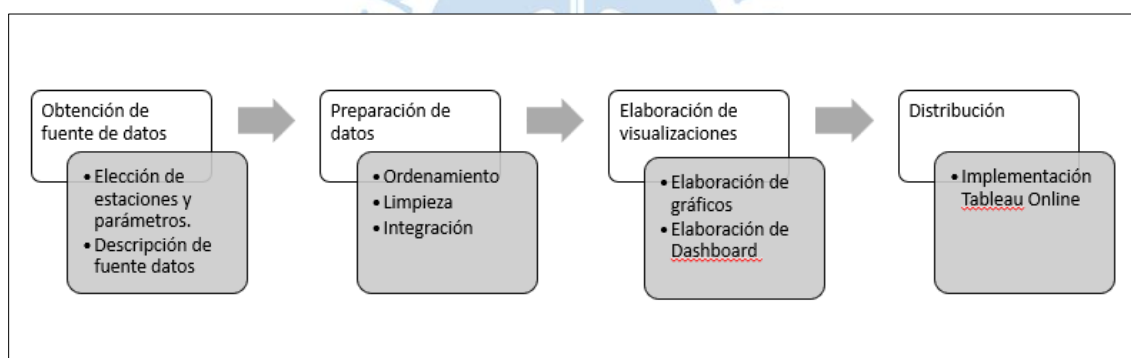


Capítulo 3

Diseño e implementación de la cadena de suministro de información

La implementación de la cadena de suministro de información se realizará a través de 4 de procedimientos, tal como lo muestra la Figura 11: (1) obtención de la fuente de datos; (2) preparación; (3) visualización; (4) distribución. Cada uno estará integrado por tareas basadas en los criterios establecidos en la etapa de metodología.

Figura 11. Procedimientos de la cadena de suministro de información



3.1 Fuente de datos

Como se analizó inicialmente, las fuentes de datos tienen diversos formatos. Por eso se desarrolló un ejecutable en Python para convertir todos los archivos en formato CSV. La finalidad de esta tarea es facilitar la importación de archivos en Tableau Prep Builder.

Luego, se procedió a trabajar en las estructuras de los archivos. La Figura 12 muestra como archivos crudos presentaban una estructura inapropiada. Específicamente, en los archivos de las estaciones de Huancabamba y San Ignacio que son de la marca NRG. Contenían información innecesaria en las primeras 169 filas, las cuales se procedieron a eliminar. Obteniendo así, una fuente de datos con menos de inconvenientes estructurales y de formato.

Figura 12. Cambio de estructuras

	A	B	C	D	E	F	G		A	B	C	D	E	F	G
141	Channel #	13						1	Date & Time Stamp	CH1Avg	CH1SD	CH1Max	CH1Min	CH2Avg	CH2SD
142	Type	1						2	1/01/2013 00:00	0.4	0.1	1.2	0.4	0.4	0
143	Description	NRG #40 Anem. m/s						3	1/01/2013 00:10	0.4	0	0.4	0.4	0.4	0
144	Details							4	1/01/2013 00:20	0.4	0	0.8	0.4	0.4	0
145	Serial Numb SN:							5	1/01/2013 00:30	0.4	0	0.4	0.4	0.4	0
146	Height	0						6	1/01/2013 00:40	0.4	0	0.4	0.4	0.4	0
147	Scale Factor	0.765						7	1/01/2013 00:50	0.4	0	0.4	0.4	0.4	0
148	Offset	0.35						8	1/01/2013 01:00	0.4	0	0.4	0.4	0.4	0
149	Units	m/s						9	1/01/2013 01:10	0.4	0	0.8	0.4	0.4	0
150								10	1/01/2013 01:20	0.4	0	0.4	0.4	0.4	0
151	Channel #	14						11	1/01/2013 01:30	0.4	0	0.4	0.4	0.4	0
152	Type	1						12	1/01/2013 01:40	0.4	0	0.4	0.4	0.4	0
153	Description	NRG #40 Anem. m/s						13	1/01/2013 01:50	0.4	0	0.4	0.4	0.4	0
154	Details							14	1/01/2013 02:00	0.4	0	0.4	0.4	0.4	0
155	Serial Numb SN:							15	1/01/2013 02:10	0.4	0	0.4	0.4	0.4	0
156	Height	0						16	1/01/2013 02:20	0.4	0	0.4	0.4	0.4	0
157	Scale Factor	0.765						17	1/01/2013 02:30	0.4	0	0.4	0.4	0.4	0
158	Offset	0.35						18	1/01/2013 02:40	0.4	0	0.8	0.4	0.4	0
159	Units	m/s						19	1/01/2013 02:50	0.4	0	0.4	0.4	0.4	0
160								20	1/01/2013 03:00	0.4	0	0.8	0.4	0.4	0
161	Channel #	15						21	1/01/2013 03:10	0.4	0	0.8	0.4	0.4	0

3.2 Preparación de datos

El procedimiento de preparación se dividirá en ordenamiento, limpieza e integración, siendo pasos secuenciales y dependientes entre sí. Todo estos se llevarán a cabo en Tableau Prep Builder.

3.2.1 Ordenamiento

El primer paso consta de la creación de flujos de integración para cada una de la fuente de datos (FD), empezando por la importación de todos los archivos de la FD correctamente estructurados. Es importante seguir la configuración inicial de importación de FD tal como se muestra en la Figura 13. De lo contrario, varias columnas se mostrarán como nulos y alterará la información original.

Figura 13. Configuración inicial

Separador de campos
Punto y coma

Calificador de texto
Automático

Conjunto de caracteres
UTF-8

Configuración regional
Español (Perú)

Siguiendo con la tarea de importación, Tableau Prep Builder, tiene la opción de unión por comodín, permitiendo importar todos los archivos guardados en una carpeta, en un solo bloque, evitando así un desorden al importar archivos de forma individual. Esta configuración,

se usó para importar todos los archivos pertenecientes a un año en un solo bloque para las estaciones de Colán y Chávez. Mientras tanto en las estaciones de UDEP, Huancabamba y San Ignacio, la herramienta facilitó la integración de todos archivos a un solo bloque dentro del flujo, debido a que aquellos contenían información de todo un año y no por meses como lo estaban las otras dos estaciones anteriores. (Ver Figura 14)

Figura 14. Integración por comodín. A) Integración por años. B) Integración por comodín.



Antes de empezar con otros pasos en el flujo, es necesario verificar y comprobar el formato de los parámetros principales, tal como se definió en la Tabla 10. Sino presentan este formato, se debe realizar el cambio, insertando un bloque de limpieza, tal como se observa en la Figura 15.

Figura 15. Cambios de formato en bloques de limpieza



Estos bloques de limpieza permiten realizar una serie de modificaciones como cambios de nombre, eliminación de columnas, campos calculados, filtraciones, entre otras. Respecto

al campo Datetime, la fecha y hora deben estar integradas en una sola columna para facilitar filtraciones y cálculos siguientes. Para el caso de los archivos de UDEP, se detallan los campos de fecha y hora de forma separada, tal como se ve en la primera imagen de la Figura 16 a. Por ende, en un bloque de limpieza se insertó un campo calculado que permita la integración de fecha y hora en una sola columna. El resultado final se muestra en Figura 16 d. Para las otras FD de las estaciones no se presentó aquel problema.

Figura 16. Unión de columnas en el campo de fecha y hora. A) Formato inicial. B) Campo calculado. C) Pasos empleados. D) Resultado final de unión de campos.

a)

Fecha	Hora
01/01/2011	2:00
01/01/2011	2:30
01/01/2011	3:00
01/01/2011	3:30
01/01/2011	4:00
01/01/2011	4:30
01/01/2011	5:00
01/01/2011	5:30
01/01/2011	6:00
01/01/2011	6:30
01/01/2011	7:00
01/01/2011	7:30

b)

Editar campo

Nombre de campo

Hora1

`STR([Fecha]) + ", " + STR(RIGHT([Hora], 8))`

c)

Cambios (14)

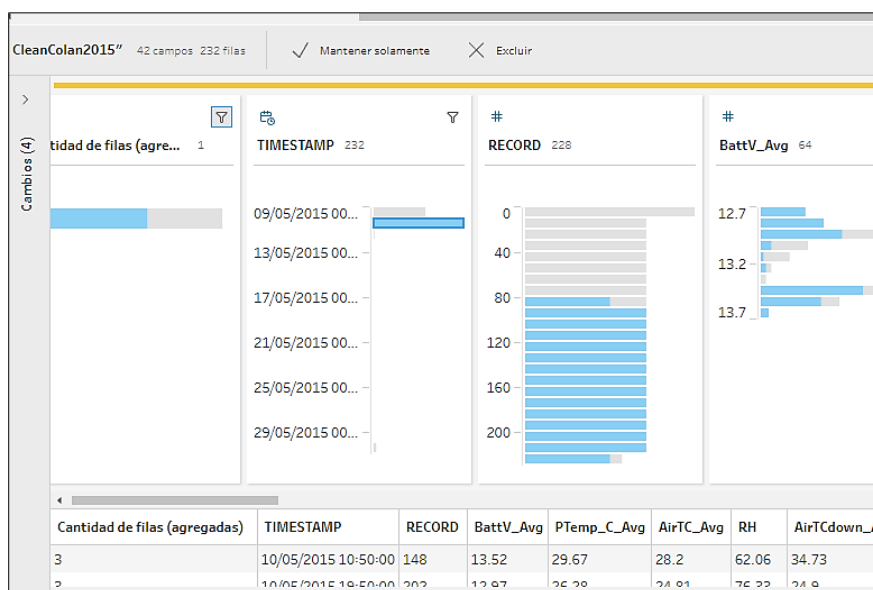
- Al tipo Fecha y hora
- Duplicar campo
 - [Fecha-1]
 - [Fecha]
- Cambiar nombre de campo
 - [Fechas_hora]
 - De [Fecha-1] a [Fechas_hora]
- Cambiar tipo
 - [Hora]
 - Al tipo Cadena
- Campo calculado
 - [Hora1]
 - `STR([Fecha]) + ", " + STR(RIGHT([Hora], 8))`
- Cambiar tipo
 - [Hora1]
 - Al tipo Fecha y hora
- Eliminar campo
 - [Fechas_hora]

d)

Hora1

01/01/2011 02:00:00
01/01/2011 02:30:00
01/01/2011 03:00:00
01/01/2011 03:30:00
01/01/2011 04:00:00
01/01/2011 04:30:00
01/01/2011 05:00:00
01/01/2011 05:30:00
01/01/2011 06:00:00
01/01/2011 06:30:00
01/01/2011 07:00:00
01/01/2011 07:30:00

Es importante determinar un formato único para las fechas debido a que, en el paso de eliminación de duplicados, se presentó un inconveniente al momento de identificar las fechas. Este inconveniente se relaciona a que las FD de las estaciones de Colán y Chávez presentaban registros con formatos de fecha distintos, algunos con mes/día/año y otro día/mes/año. Tableau los interpretaba de igual manera, como se ve en la Figura 17. Por tal motivo se determinó el formato para todos los archivos día/mes/año.

Figura 17. Formato de fechas distintas

Luego de definir el formato de fechas, se procedió a eliminar datos duplicados. En primer lugar, se realizó un mapeo del rango de registros abarcado por cada archivo, con la finalidad de observar la cantidad de repeticiones existentes.

La Figura 18 muestra un análisis de registros repetidos para Colán, siendo el bloque del año 2019 con más datos duplicados. La Figura 19 muestra el número de registros repetidos para Chávez, donde el bloque 2020 presenta mayor número de registros repetidos. La FD de las otras estaciones no presentan este problema. Estos análisis dan a conocer la dificultad de limpieza de duplicados si se realizara de manera manual en hojas Excel.

Figura 18. Matriz de datos repetidos en la Estación Colán

Archivos de flujo Colán			
Archivos	Intervalo de tiempo		N° registros repetidos
	Inicio	Fin	
Colan2013	1/01/2013	31/12/2013	0
Colan2014	1/01/2014	31/12/2014	0
Colan2015	29/08/2014	30/12/2015	17856
Colan2016	1/12/2015	28/12/2016	4176
Colan2017	1/12/2015	13/10/2017	56592
Colan2018	13/10/2017	23/07/2018	0
Colan2019	7/05/2016	16/10/2019	116208
Colan2020	1/01/2020	1/12/2020	11088
Colan2021	1/12/2020	7/01/2022	0

Figura 19. Matriz de datos repetidos en la Estación Chávez

Archivos de flujo Chávez			
Archivos	Intervalo de tiempo		N° de registros
	Inicio	Fin	
Chavez2010	23/05/2010	30/11/2010	2808
Chavez2011	1/12/2010	2/11/2011	984
Chavez2012	1/12/2011	3/12/2012	624
Chavez2013	1/12/2012	30/11/2013	2088
Chavez2014	1/12/2013	30/11/2014	5112
Chavez2015	1/12/2014	18/12/2015	8784
Chavez2016	24/05/2016	23/12/2016	5256
Chavez2017	24/05/2016	10/12/2017	11184
Chavez2018	6/05/2017	3/09/2018	31392
Chavez2019	9/02/2018	1/06/2019	81648
Chavez2020	23/03/2019	13/12/2020	62640
Chavez2021	13/12/2020	15/10/2021	0

Retomando al flujo de Tableau Prep Builder, se integra un paso denominado “Agregar” para eliminar duplicados, siempre y cuando se coloque la condición de que los valores de todas las columnas sean iguales para que la herramienta pueda agruparlos a una sola fila y elimine los archivos sobrantes, logrando así una solución eficiente para esta dificultad sin importar el número de repeticiones que tenga. (Ver Figura 20)

Figura 20. Eliminación de duplicados con paso "Agregar"

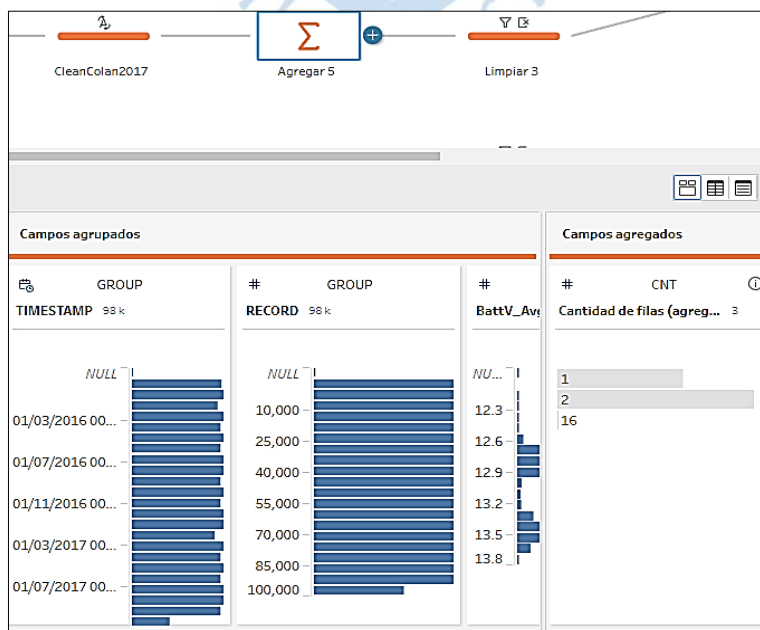


Figura 21. Paso de "Salida" en flujo UDEP

Salida

8 campos

Guardar salida en

Archivo

Ejemplar

Nombre

UDEP_Integrado

Ubicación

C:\Users\Jefferson\Documents\Universidad\Tesis\Tesis1\Flujos de Integración

Tipo de salida

Valores separados por comas (.csv)

Escribir opciones

Seleccione una opción para crear o actualizar la tabla de salida.

Actualización completa

Crear tabla

Ejecutar flujo

Guardar en UDEP_Integrado.csv

DATATIME	Temp_air	RH_%	Wind_Speed	Wind_Dir	Presion_Bar	Rain_mm	Radiation_Solar
17/05/2020 08:20:00	21.5	79	1.3	NNE	1,008.9	0	373
18/09/2018 15:40:00	27.4	55	2.2	NNE	1,003.1	0	550
07/12/2021 14:40:00	30.6	50	1.8	NNE	1,003.7	0	633
13/05/2017 08:20:00	22.9	81	0	NNE	1,010	0	161
26/03/2020 11:00:00	30.7	51	0.9	NNE	1,006.8	0	717
15/10/2020 09:30:00	21.9	74	1.3	WSW	1,008.7	0	652
13/08/2021 19:00:00	21.6	74	1.3	NNE	1,010.3	0	0
20/04/2019 17:30:00	29.9	52	3.6	NNE	1,004.7	0	147
27/12/2016 22:00:00	24.2	71	0	NNE	1,007.3	0	0
11/04/2011 04:00:00	20.3	87	0.4	SSE	1,008	0	0
15/01/2019 03:30:00	23.9	75	0	---	1,006.6	0	0
04/12/2021 07:50:00	20.1	82	0.4	SE	1,008.3	0	54
17/07/2021 12:00:00	24.2	69	1.3	NNE	1,008.8	0	536
14/10/2019 13:10:00	29.1	50	1.3	NNE	1,006.9	0	961
28/08/2017 23:20:00	18.7	86	0	NNE	1,010.4	0	0

Este procedimiento se hizo en dos pasos generales, primero, se realizó una previsualización de los datos ordenados en la etapa pasada mediante Tableau Desktop, con la finalidad de observar atípicos y en qué rango de tiempo se encuentran. Se realizaron gráficas de forma general para cada parámetro y estación meteorológica.

En Colán se encontraron anomalías para la radiación solar, mostrando datos negativos y más de 5000 w/m² para el año 2014. De igual manera sucedía con el viento, indicando valores por más de 100 m/s y respecto a orientación, presentaban valores fuera de los 360 grados sexagesimales. Finalmente, en presión atmosférica, señalaba valores por menos de 600 mBar. Todas estas anomalías se muestran en la Figura 22.

En la estación Chávez, se encontraron anomalías para temperaturas, mostrando valores en el año 2017 por más de 50 °C. En radiación solar, arrojaba valores negativos totalmente errados y, finalmente, en lluvia acumulada en el año 2017, registró más de 13000 mm, valor totalmente irreal. Una posible causa es el fallo de los sensores y *data logger* debido al evento climático Niño Costero. (Ver Figura 23)

Figura 22. Anomalías en la estación Colán. A) Anomalías Radiación Solar. B) Anomalías velocidad de viento. C) Anomalías Dirección del viento. D) Anomalías Presión atmosférica.

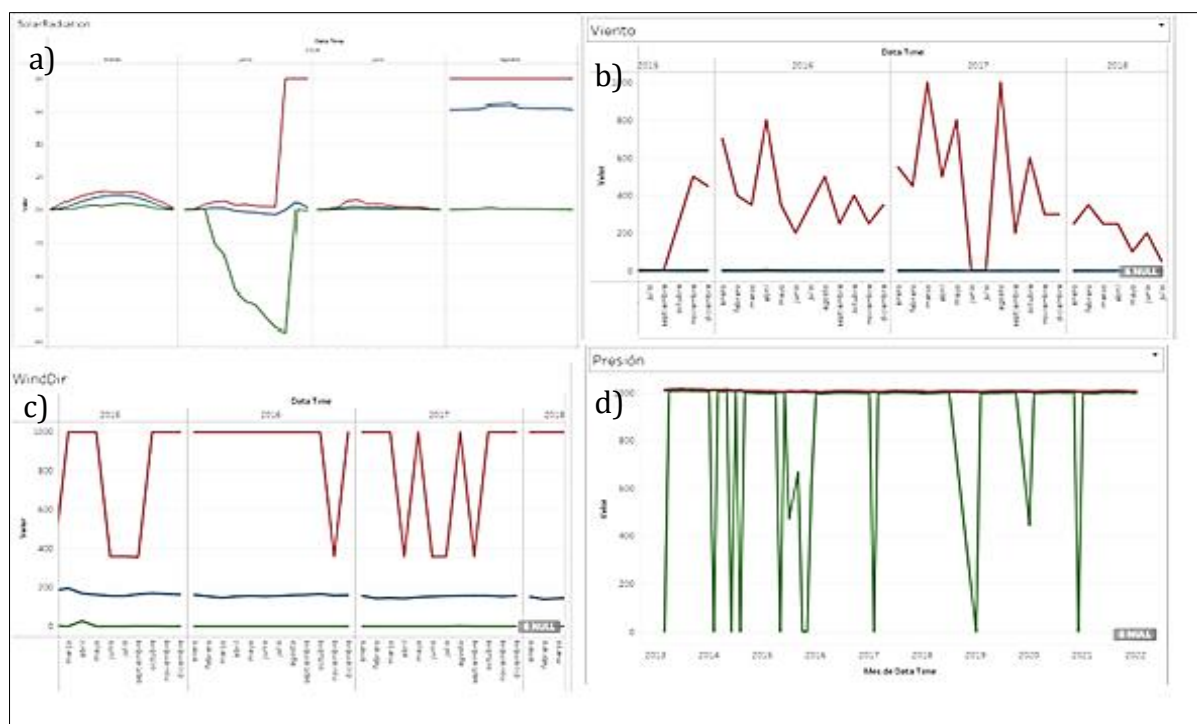
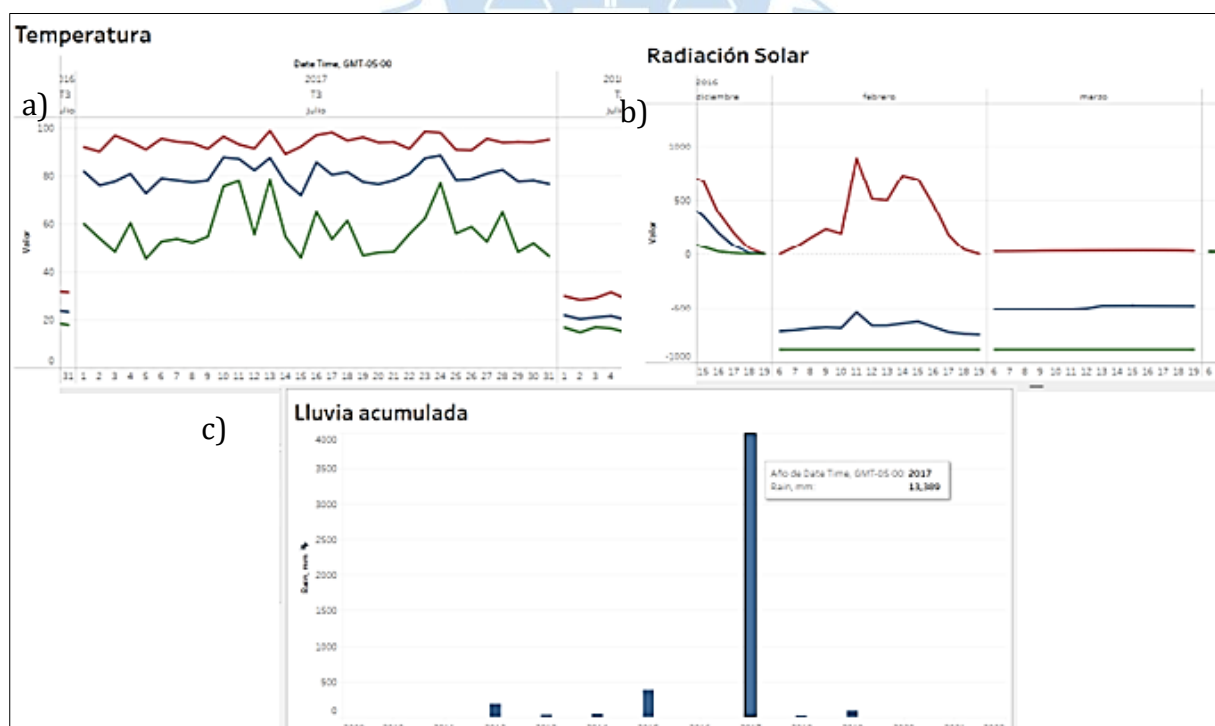


Figura 23. Anomalías en la estación Chávez. A) Anomalías temperatura. B) Anomalías Radiación Solar. C) Anomalías Lluvia acumulada.



Para el resto de las estaciones no se presentaron rangos de anomalías considerables, sin embargo, si existían algunas de ellas.

Luego de realizar una previsualización, se trabajaron nuevos flujos en Tableau Prep Builder para cada estación con el objetivo de limpiar los datos atípicos. Para cada uno de los flujos, se agregó un bloque de limpieza en donde se insertó un campo calculado utilizando código y funciones proporcionadas por Tableau con el fin de convertir todos los datos atípicos a *null* según las características establecidas en la Tabla 11.

En la Figura 24 muestran las funciones insertadas en los campos calculados para temperaturas. Para las tres primeras estaciones no varía; sin embargo, para Huancabamba y San Ignacio, al poseer otra geografía, los parámetros de temperatura llegan a cambiar. (Ver Figura 25)

Figura 24. Funciones para limpieza de datos temperatura

Editar campo	
Nombre de campo	
Temp_air	
<pre>IIF([Temp, °C 1] > 40, NULL, IIF([Temp, °C 1] < 10, NULL, [Temp, °C 1]))</pre>	

Figura 25. Funciones para limpieza de temperatura (Huancabamba y San Ignacio)

Editar campo	
Nombre de campo	
Temp_air	
<pre>IIF([Temp, °C 1] > 40, NULL, IIF([Temp, °C 1] < 8, NULL, [Temp, °C 1]))</pre>	

Para humedad relativa, se insertó una función general para todos. Si es menor a 30 y mayor a 100, estos datos se transformarán a nulos. (Ver Figura 26)

Figura 26. Función para limpieza de los datos de Humedad

Editar campo	
Nombre de campo	
RH_%	
<pre>IF [RH_\$()] > 100 Then NULL ELSEIF [RH_\$()] < 30 THEN NULL else [RH_\$()] End</pre>	

Para dirección del viento, todos los datos deben estar dentro del rango de 0 a 360°. Por ende, la misma función de la Figura 27 se aplicó para todas las estaciones.

Figura 27. Función de limpieza para la dirección del viento

Editar campo	
Nombre de campo	Wind_Dir
	<code>IIF([WindDir]>360, NULL, [WindDir])</code>

Para velocidad del viento, se aplicará la misma función para todos porque los valores no pueden superar los 20 m/s para todas las estaciones. (Ver Figura 28)

Figura 28. Función de limpieza para la velocidad del viento

Editar campo	
Nombre de campo	Wind_Speed
	<code>IIF([WS_ms_Avg]>20,NULL,[WS_ms_Avg])</code>

Para radiación solar, los valores deben encontrarse dentro de un rango de 0 a 1200 w/m². Este criterio es igual para todas las estaciones. Dentro de este se debe tener presente una condición más, los valores mayores a 1200 y menores a 2000, serán considerados como un máximo de 1200. (Ver Figura 29)

Figura 29. Funciones para limpieza de parámetro de radiación solar

Editar campo	
Nombre de campo	Solar_Radiation
	<pre> If [SR01Up_Avg] < -100 OR [SR01Up_Avg] > 2000 THEN NULL ELSEIF [SR01Up_Avg]>= -100 and [SR01Up_Avg]< 0 THEN 0 ELSEIF [SR01Up_Avg] > 1200 and [SR01Up_Avg] <= 2000 Then 1200 ELSE [SR01Up_Avg] END </pre>

En presión atmosférica, sí se emplean rangos de valores distintos por la geografía variada que hay entre las estaciones, menos en Colán y UDEP, ya que son similares. (Ver Figura 30)

Figura 30. Funciones para limpieza de presión atmosférica.

The figure displays three screenshots of a code editor interface, each showing a different set of conditional logic for cleaning atmospheric pressure data. Each screenshot has a title bar that says 'Editar campo' and a text input field for the field name.

- Top Left Screenshot:** The field name is 'Presion_Bar'. The code is:


```
IF [Presion_Bar()] < 1000
then NULL
ELSEIF [Presion_Bar()] > 1120
then Null
Else [Presion_Bar()]
end
```
- Top Right Screenshot:** The field name is 'Presion_Bar1'. The code is:


```
// Huancabamba
IF [Presion_Bar] < 790
then NULL
ELSEIF [Presion_Bar] > 810
THEN NULL
end
```
- Bottom Screenshot:** The field name is 'Presion_Bar1'. The code is:


```
// Huancabamba
IF [Presion_Bar] < 790
then NULL
ELSEIF [Presion_Bar] > 810
THEN NULL
end
```

Finalmente, para el parámetro de lluvia acumulada en Chávez para el año 2017, se eliminaron todos los datos, dado que estos se encontraban totalmente fuera de un rango normal, debido a que, durante el fenómeno del Niño Costero, los sensores se dañaron.

Siguiendo en el flujo de limpieza, se procedió a aplicar el procedimiento de remplazo de datos nulos. Como se determinó en metodología, se reemplazarían a través del promedio de las dos posiciones previas y posteriores del dato nulo. Dentro de este procedimiento, se determinó primero una lógica en Excel aplicando distintas funciones que permitan cumplir con este razonamiento.

Comprobada una vez el funcionamiento de la lógica anterior, se desarrolló un algoritmo en el lenguaje de Python dentro de la plataforma Colab¹ con la finalidad de confirmar si no hay algún inconveniente en su ejecución. Posteriormente, se copia el algoritmo en una script de Python y se guarda dentro del ordenador. La lógica de este algoritmo se aplicó de igual manera para las estaciones de Colán, Udep, Huancabamba y San Ignacio. Sin embargo,

¹ Plataformar de google para poder insertar algoritmos de lenguajes de programación. <https://colab.research.google.com/?hl=es>

para la de Chávez se obvió la inserción de los parámetros de orientación del viento y presión atmosférica porque no se registran. (Ver Figura 31)

Figura 31. Algoritmo Python empleado en datos nulos

```

1  import numpy as np
2
3
4  def get_output_schema():
5      return pd.DataFrame({
6          'Temp_Air': prep_decimal(),
7          'Wind_Speed': prep_decimal(),
8          'Wind_Dir': prep_int(),
9          'Solar_Radiation': prep_decimal(),
10         'DATETIME': prep_datetime(),
11         'RH_%': prep_decimal(),
12         'Presion_Bar': prep_decimal(),
13         'Rain_mm': prep_decimal(),
14         'RH_X1': prep_decimal(),
15         'Presion_Bar1': prep_decimal(),
16     })
17
18
19
20 def funcion(df):
21     df = df.sort_values("DATETIME")
22     df.reset_index(inplace=True, drop=True)
23
24     headers = ["RH_%", "Presion_Bar"]
25
26     for header in headers:
27
28         df[header+"1"] = df[header]
29
30         indices_de_NaN = df[df[header].isna()].index.to_list()
31
32         values_to_replace = []
33         indices_de_NaN2 = indices_de_NaN.copy()
34
35         for indice in indices_de_NaN:
36             try:
37                 tmp = np.array([df[header][indice-1], df[header][indice-2], df[header][indice+1], df[header][indice+2]])
38                 values_to_replace.append(np.nanmean(tmp))
39             except:
40                 indices_de_NaN2.pop(indice)
41                 continue
42
43         for c, indice in enumerate(indices_de_NaN2):
44             df[header+"1"][indice] = values_to_replace[c]
45
46     return df
47
48

```

Para aplicar esto en el flujo de Tableau Prep Builder, en primer lugar, se descarga la extensión de Tabpy, la cual permitirá poder agregar un paso denominado script en el flujo y se pueda ejecutar correctamente. En las configuraciones se agrega la ubicación guardada del script y se le añade el nombre de la función. (Ver Figura 32)

Figura 32. Configuración Tabpy

Script San Ignacio 0 campos Filtrar valores...

Configuración Cambios (0)

Tipo de conexión

☐ Rserve

☒ Servidor Python (TabPy) de Tableau

Servidor

Conexión a localhost:9004

Nombre de archivo

archivo_flujo.SanIgnacio.py

Examinar

Nombre de función

funcion

Incluya la función de esquema "get_output_schema" en su script para definir los campos y tipos de datos devueltos. [Más información](#)

Concluida la ejecución de Tabpy, se agrega un bloque de limpieza donde se eliminarán los campos anteriores y reemplazan los nuevos con el nombre correspondiente. Finalmente, se agrega un paso de salida para poder exportar los nuevos datos limpios. Estos de igual manera deben estar en formato CSV.

3.2.3 Integración

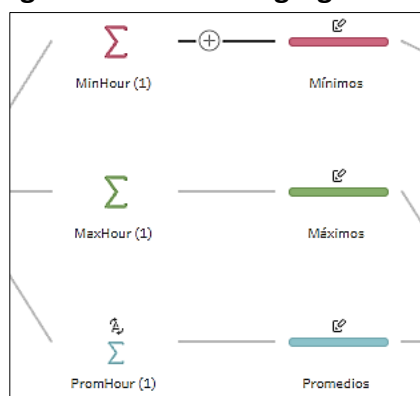
En esta última etapa de preparación, se crea un nuevo flujo nuevamente para cada estación. Una vez importados los datos de la etapa anterior, se agregaron bloques de limpieza con la finalidad de aplicar campos calculados para separar la hora y fecha del campo "Datetime". Luego estos se integraron a una nueva columna denominada "DateHour". Dado que, se realizará una agrupación horaria, es decir que todos los datos que presenten la misma hora y fecha se integrarán, tal como se definió. Cabe precisar, que si se desea realizar otra integración ya sea de formar diaria, mensual o anual, se aplica la misma lógica. (Ver Figura 33)

Figura 33. Función para separar las fecha y hora

Editar campo	Editar campo
Nombre de campo	Nombre de campo
DATE	Hour
<code>DATE (LEFT (STR ([DATETIME]), 10))</code>	<code>DATEPART ("hour", [DATETIME])</code>

Luego, se agregaron tres bloques de agregar de forma paralela, tal como se observa en la Figura 34. Este paso va a permitir que, al aplicar un criterio de agrupación. Para este caso, todos los datos que tengan la misma hora y fecha se agruparán obteniendo datos máximos, promedios y mínimos según los criterios de agregación establecidos en la Tabla 12 para cada parámetro. La implementación de este procedimiento optimiza el tiempo de agrupación, en comparación a su implementación con funciones en Excel.

Figura 34. Pasos de agregación

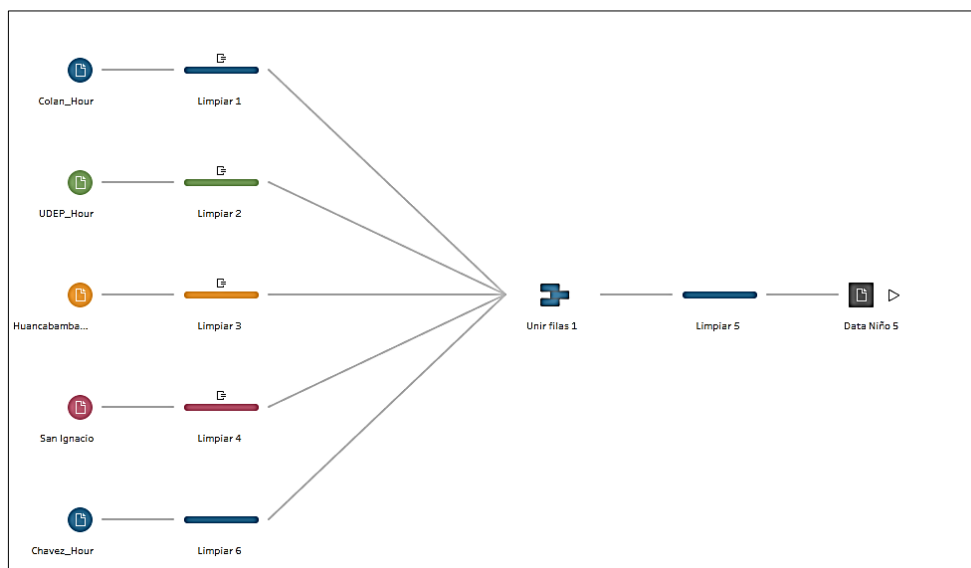


Para finalizar este flujo, se añade el paso de “Salida” para exportar los datos mediante el formato CSV.

Teniendo la data limpia e integrada con todos los criterios establecidos, se procede a crear un nuevo flujo, donde se integren los archivos finales de la etapa anterior de cada una de las estaciones. Para cada archivo, en un paso de limpieza se añaden los campos con la información de cada estación: latitud, longitud, ubicación, provincia y departamento, como se observa en la Figura 36. Esto, con el fin de poder identificarlas al momento de integrar todas las visualizaciones. Finalmente, se agrega un paso de salida y se exportan mediante el formato CSV, concluyendo así con el último flujo de preparación de datos. (Ver Figura 37)

Figura 36. Campos de ubicación

<div> <div>Limpiar 3</div> <div>20 campos 79K filas</div> <div>Filtrar valores...</div> <div>Cambiar el nombre de los campos...</div> <div>Crear campo calculado...</div> </div>				
Cambios (1)	<div> <div>Abc</div> <div>Ubicación 1</div> <div>Huancabamba</div> </div>	<div> <div>Abc</div> <div>Region 1</div> <div>Piura</div> </div>	<div> <div>Abc</div> <div>Provincia 1</div> <div>Huancabamba</div> </div>	<div> <div>#</div> <div>Latitud 1</div> <div>-5.2366667</div> </div>
				<div> <div>#</div> <div>Longitud 1</div> <div>-79.448333333</div> </div>

Figura 37. Flujo final de preparación

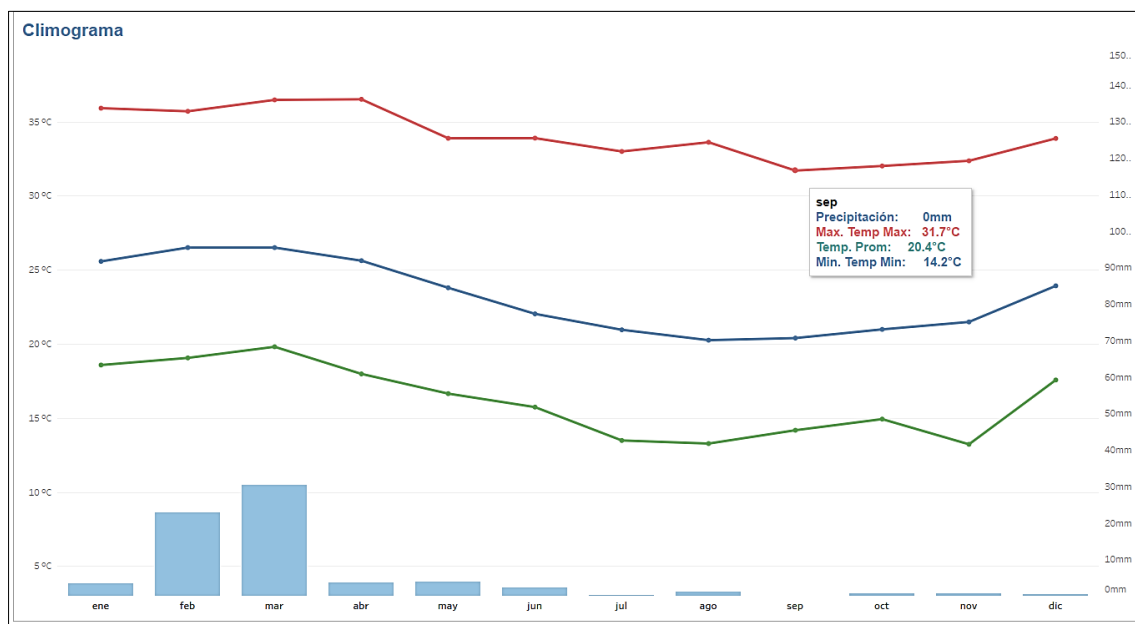
3.3 Visualizaciones

La implementación de visualizaciones se trabajó en Tableau Desktop. En primer lugar, se importó la data final del procedimiento anterior. Luego, se crearon los gráficos con sus respectivos criterios establecidos en la Tabla 12 para cada estación meteorológica. Además, las visualizaciones se categorizaron de dos formas:

A. Visualizaciones de resumen mensual

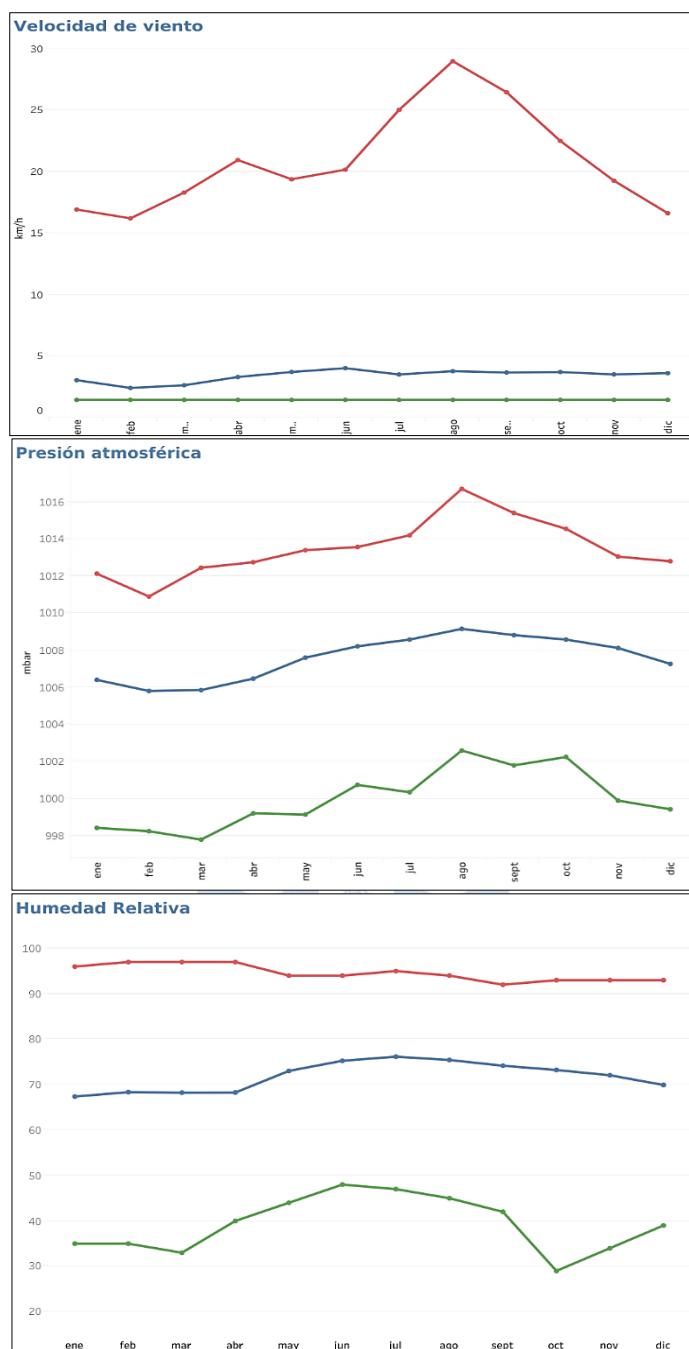
Se crearon 8 gráficos para cada estación meteorológica, a excepción de la de Chávez, ya que no registra los parámetros de presión y dirección del viento, siendo en total 38 gráficos. Estos contienen información mensual de un total aproximado de 7 años. Además, tienen la facilidad de seleccionar uno o más rango de años a comparar. Todos los gráficos poseen una etiqueta, permitiendo al usuario colar el cursor en un punto del gráfico y observar la información para ese periodo del tiempo. Siendo totalmente intuitivo al momento de visualizar la información que se desea.

Los climogramas están seccionados en temperaturas máximas, promedio y mínima, adicionalmente, poseen un gráfico de barras haciendo referencia a la cantidad de lluvia acumulada promedio mensual. (Ver Figura 38)

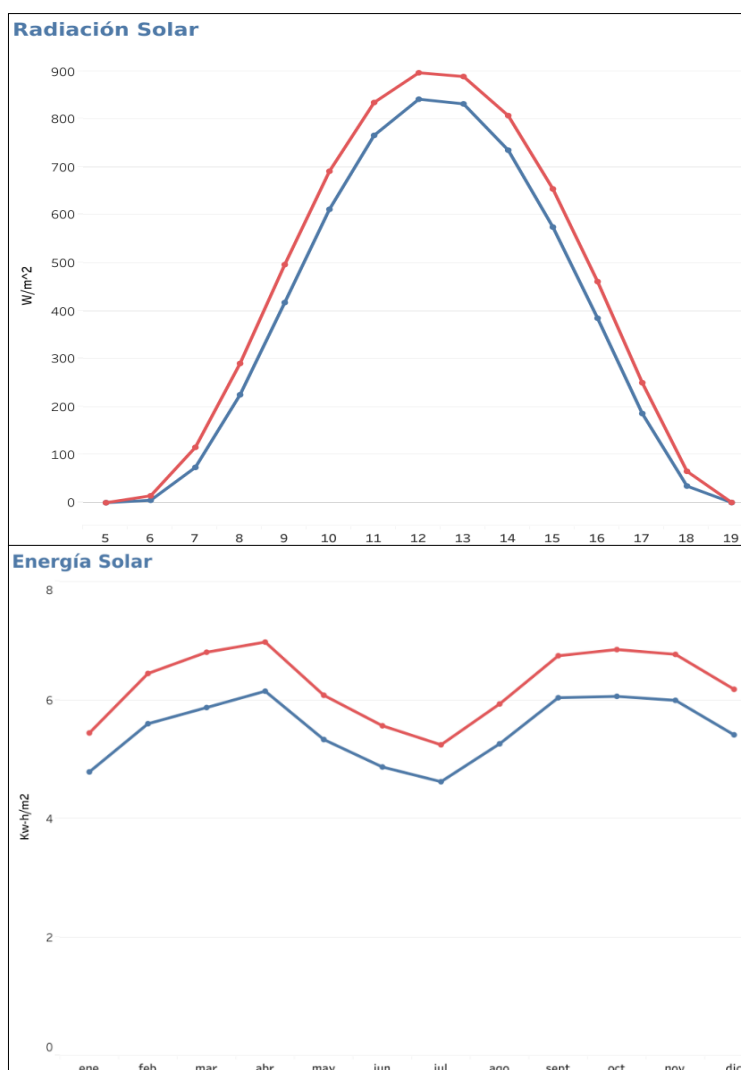
Figura 38. Ejemplo de climograma

En el gráfico de líneas de la Humedad relativa, presión atmosférica y velocidad del viento presentan las mismas características respecto a los parámetros de máximos, promedios y mínimos. Cada una se distingue por un color en específico rojo, azul y verde respectivamente. (Ver Figura 39)

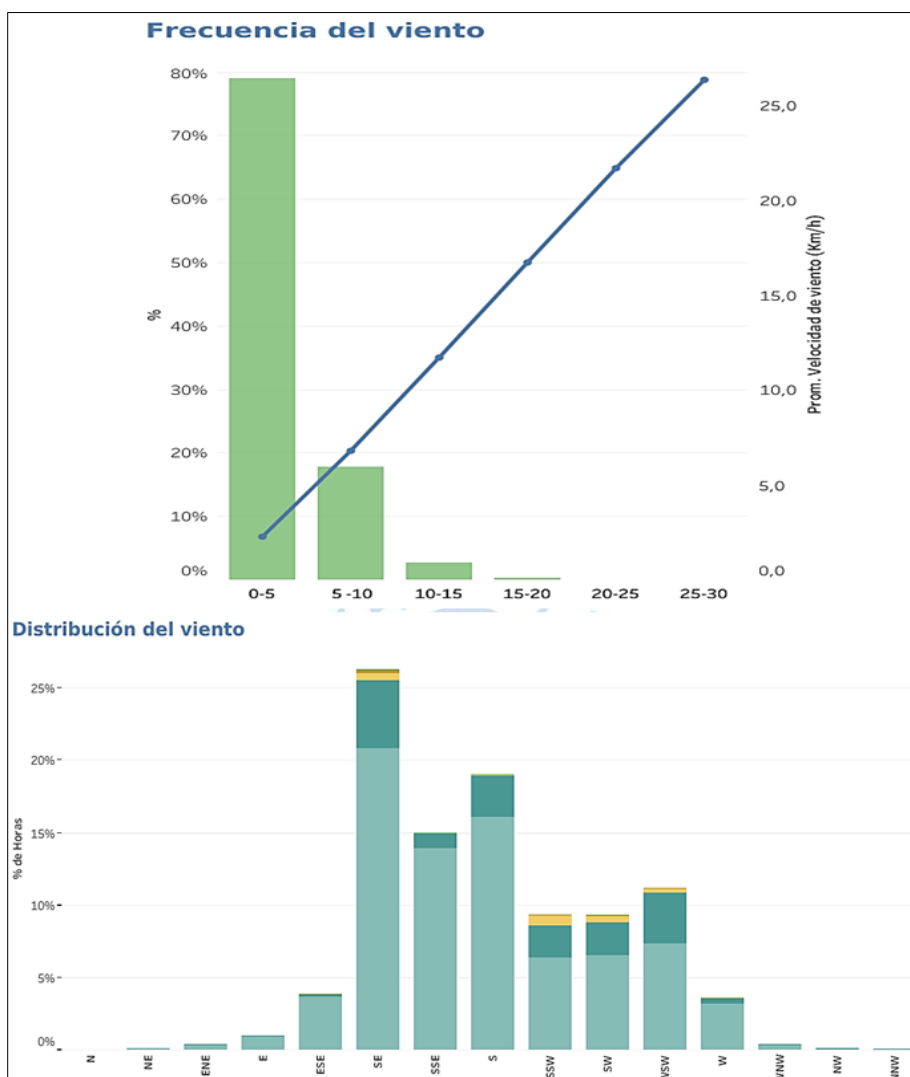
Figura 39. Gráficos de Visualización de velocidad de viento, humedad relativa y presión atmosférica.



Con referencia al parámetro de radiación solar se crearon dos gráficos, una que indica este parámetro de forma horario, mostrando las horas solar pico promedio que hay durante el día y luego el de energía solar, el cual, muestra el potencial existente local. De igual manera, estos se caracterizan por máximo y promedio. (Ver Figura 40)

Figura 40. Visualizaciones de radiación y energía solar.

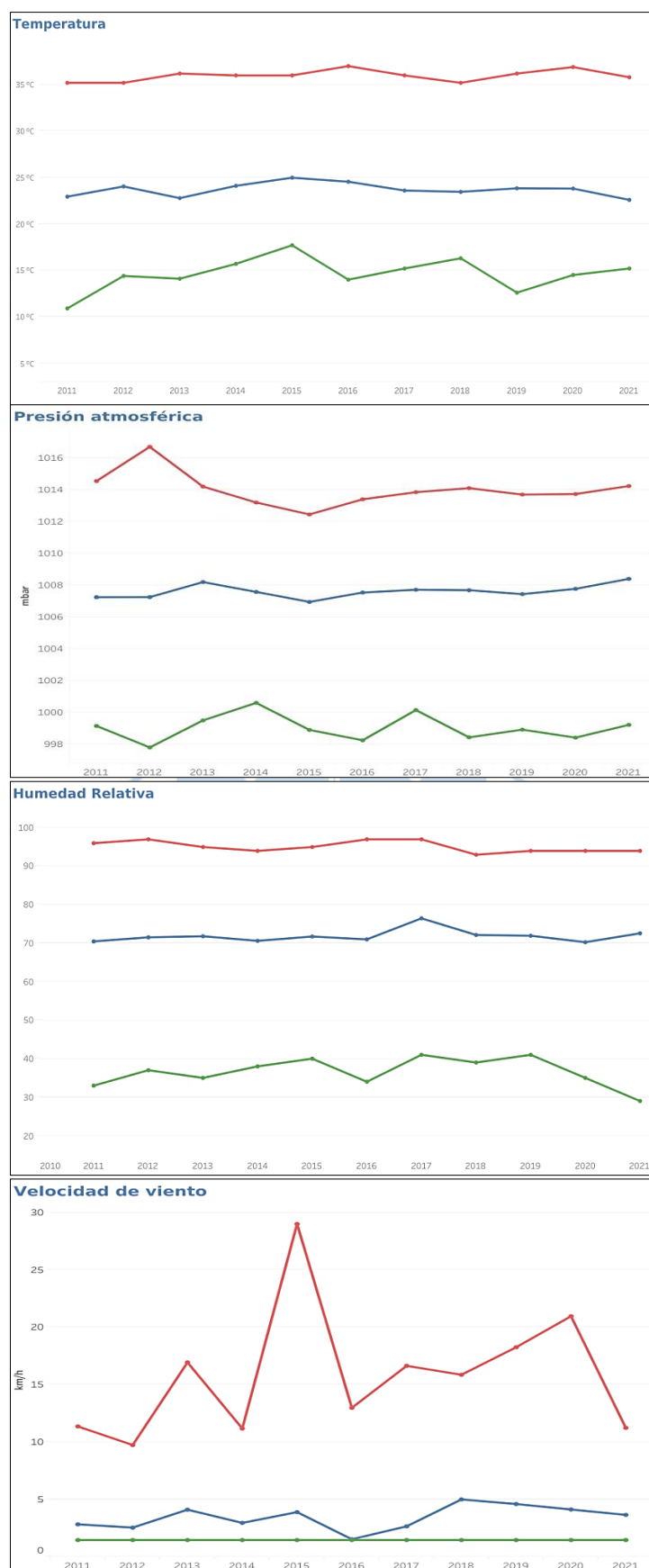
Finalmente, con la orientación del viento se crearon dos gráficas. Una de estas es la distribución donde señala como se distribuye el viento en base al rumbo y velocidad de este, determinado por porcentajes. Además, se muestra la frecuencia de vientos, con la finalidad de observar qué rangos de velocidad son más frecuentes junto a su velocidad promedio por rangos. (Ver Figura 41)

Figura 41. Visualizaciones para frecuencia y distribución del viento.**B. Visualizaciones de resumen anual**

De igual manera se crearon 8 gráficos para cada FD preparada de las estaciones meteorológicas. Siendo un total de 38 visualizaciones, con la opción de filtrar por años y meses, según el requerimiento del usuario. Sin embargo, para este caso no aplicará el gráfico de radiación solar horaria y será reemplazada por el total de lluvia acumulada durante los años y/o meses filtrados.

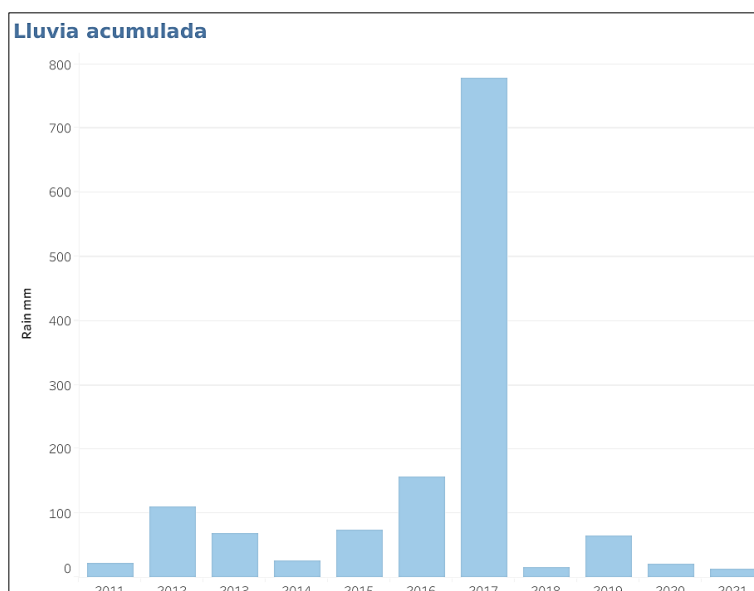
La Figura 42 grafica los valores máximos y mínimos absolutos junto al promedio por año de todos los meses del año. Se repite la lógica del uso de colores rojo, azul y verde para los máximos, promedios, y mínimos respectivamente.

Figura 42. Visualizaciones de temperatura, presión, humedad y velocidad del viento anual.



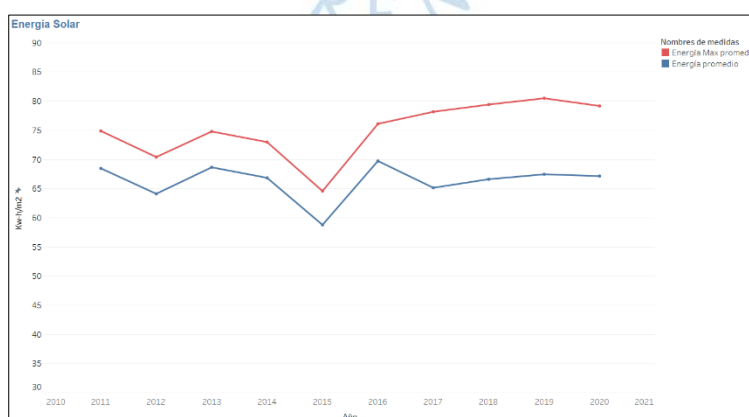
Las visualizaciones de lluvia acumulada se ven reflejada en la misma lógica utilizada en un climograma, en donde a través de barras se puede visualizar el valor acumulado por años. Entrando un poco en análisis, el año 2017 resaltó por ser el Fenómeno de El Niño costero, lo cual trajo como consecuencia inundaciones por las continuas precipitaciones. (Ver Figura 43)

Figura 43. Visualización de lluvia acumulada anual.



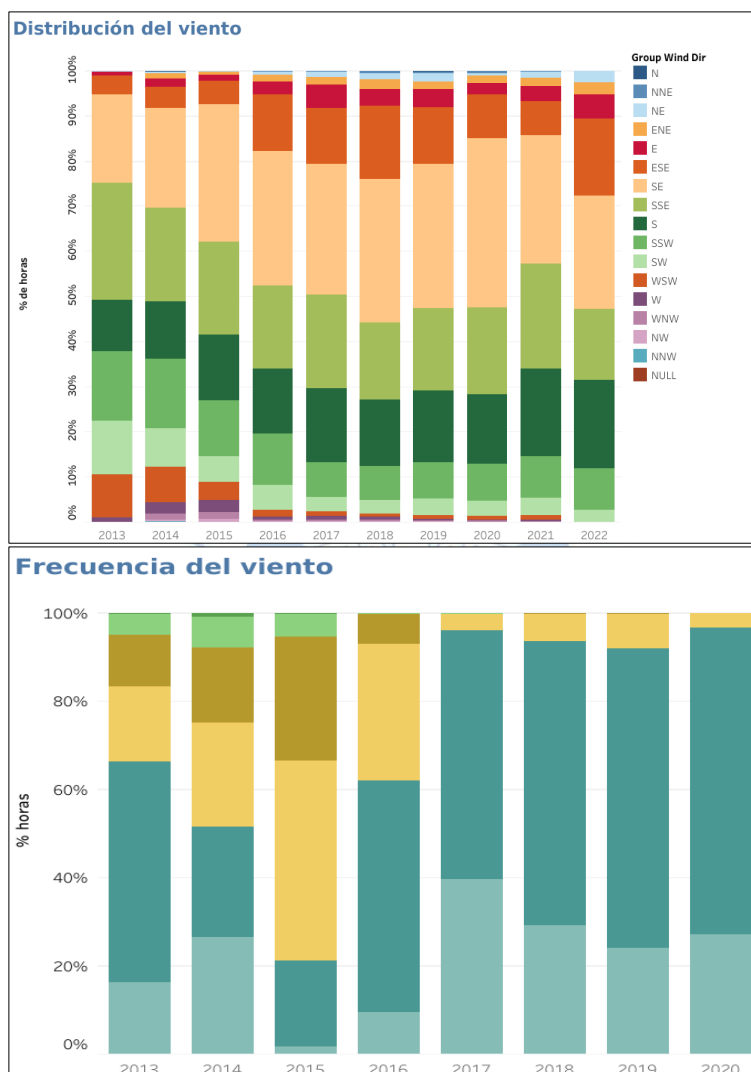
Respecto al gráfico de energía acumulada, se consideró la suma de los valores por año y convertidos a Kw-h. De igual manera, el máximo y promedio se ve considerada por los colores rojo y azul respectivamente. Estos gráficos indicarán el comportamiento del recurso solar por años y meses filtrados, siendo estos estudios preliminares para la instalación de próximos proyectos de energía solar. (Ver Figura 44)

Figura 44. Energía acumulada anualmente



Finalmente, las visualizaciones respecto al comportamiento de viento se categorizan en orientación y frecuencia. En donde por años y meses filtrados, mostrará el porcentaje de horas de frecuencia según sea su rumbo y velocidad. Estos gráficos ayudan a verificar el rumbo de la velocidad de viento para estudios preliminares de instalación de proyectos eólicos. (Ver Figura 45)

Figura 45. Distribución y frecuencia de vientos anuales



Campos calculados

En este procedimiento gráfico también se insertaron campos calculados para facilitar la creación de las visualizaciones de energía solar, distribución y frecuencias de vientos.

El gráfico de energía solar deriva de los datos de radiación. Para obtener la energía se aplica una sumatoria de todos estos datos divididos por la cantidad de días que hay en un año. Por ejemplo, si el usuario necesita saber la energía promedio del mes de marzo de 3 años, el algoritmo obtendrá un promedio de toda la suma de los datos de radiación solar para esa

cantidad de años en el mes escogido. La misma lógica se aplica para energía solar máxima. Las unidades de energía se detallan en Kw/h por tal razón se divide entre 1000. (Ver Figura 46)

Figura 46. Algoritmos para energía solar promedio y máxima.

The image shows two Excel formula bars. The top one is for 'Energía promedio' and the bottom one is for 'Energía Max promedio'. Both formulas calculate the average of solar radiation over a specific period, divided by 1000.

Energía promedio

$$\frac{\text{SUM}([\text{Prom Solar Radiation}])}{(\text{COUNTD}(\text{DAY}([\text{Date Hour}])) * \text{COUNTD}(\text{YEAR}([\text{Date Hour}])))} / 1000$$

Energía Max promedio

$$\frac{\text{SUM}([\text{Max Solar Radiation}])}{(\text{COUNTD}(\text{DAY}([\text{Date Hour}])) * \text{COUNTD}(\text{YEAR}([\text{Date Hour}])))} / 1000$$

Respecto a los gráficos de distribución y frecuencia de viento, como primer paso las unidades de velocidad se transformaron a Km/h aplicando la conversión correspondiente. Luego, para la agrupación de las direcciones se aplicó el criterio de rangos de la Tabla 7 como se detalla en la Figura 47.

Figura 47. Algoritmo para agrupación de dirección del viento.

The image shows an Excel formula bar for 'Group Wind Dir'. The formula uses a series of IF-THEN-ELSE statements to categorize wind directions based on specific ranges of values.

```

IF [Prom Wind Dir] >11.25 and [Prom Wind Dir] <= 33.75
Then "NNE"
ELSEIF [Prom Wind Dir] >33.75 and [Prom Wind Dir] <= 56.25
Then "NE"
ELSEIF [Prom Wind Dir] > 56.25 and [Prom Wind Dir] <= 78.75
Then "ENE"
ELSEIF [Prom Wind Dir] > 78.75 and [Prom Wind Dir] <= 101.25
Then "E"
ELSEIF [Prom Wind Dir] > 101.25 and [Prom Wind Dir] <= 123.75
Then "ESE"
ELSEIF [Prom Wind Dir] > 123.75 and [Prom Wind Dir] <= 146.25
Then "SE"
ELSEIF [Prom Wind Dir] > 146.25 and [Prom Wind Dir] <= 168.75
Then "SSE"
ELSEIF [Prom Wind Dir] > 168.75 and [Prom Wind Dir] <= 191.25
Then "S"
ELSEIF [Prom Wind Dir] > 191.25 and [Prom Wind Dir] <= 213.75
Then "SSW"
ELSEIF [Prom Wind Dir] > 213.75 and [Prom Wind Dir] <= 236.25
Then "SW"
ELSEIF [Prom Wind Dir] > 236.25 and [Prom Wind Dir] <= 258.75
Then "WSW"
ELSEIF [Prom Wind Dir] > 258.75 and [Prom Wind Dir] <= 281.25
Then "W"
ELSEIF [Prom Wind Dir] > 281.25 and [Prom Wind Dir] <= 303.75
Then "WNW"
ELSEIF [Prom Wind Dir] > 303.75 and [Prom Wind Dir] <= 326.25
Then "NW"
ELSEIF [Prom Wind Dir] > 326.25 and [Prom Wind Dir] <= 348.75
Then "NNW"
ELSEIF [Prom Wind Dir] > 348.75 and [Prom Wind Dir] <= 359
Then "N"
ELSEIF [Prom Wind Dir] > 359 and [Prom Wind Dir] <= 11.25
Then "N"
ELSEIF [Prom Wind Dir] = NULL
Then "-"
END
  
```

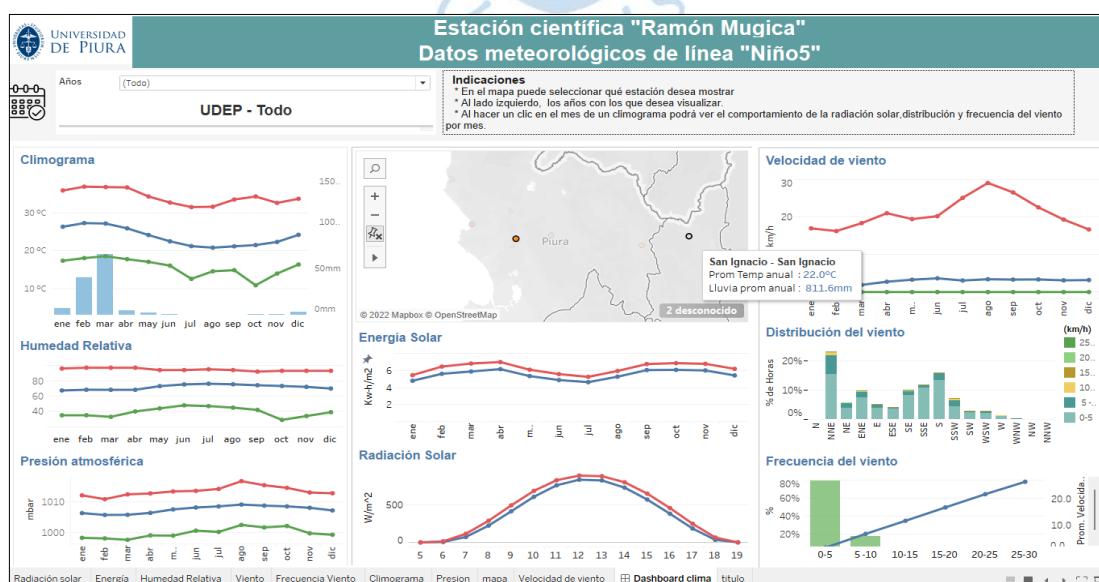
Implementación de Dashboard

El diseño y elaboración de los 2 *dashboard* parte con la integración de todas las visualizaciones creadas anteriormente, sin embargo, éstas deben mostrarse de una manera ordenada con el objetivo de dar a entender la información. Estos *dashboard* permiten al usuario tener una interactividad con la información que desea ver, puesto que, puede filtrarla por años (2011 hasta 2021), por estación y por meses, donde las 5 ubicaciones se encuentran posicionadas en el mapa las cuales son activadas mediante un clic. Como se puede observar en Figura 48. Adicionalmente, se le añadieron indicaciones para facilitar una guía de interacción para el usuario concluyendo así con la elaboración de los *dashboard* como se observa la Figura 49² y Figura 50.³

Figura 48. Filtros de dashboard.



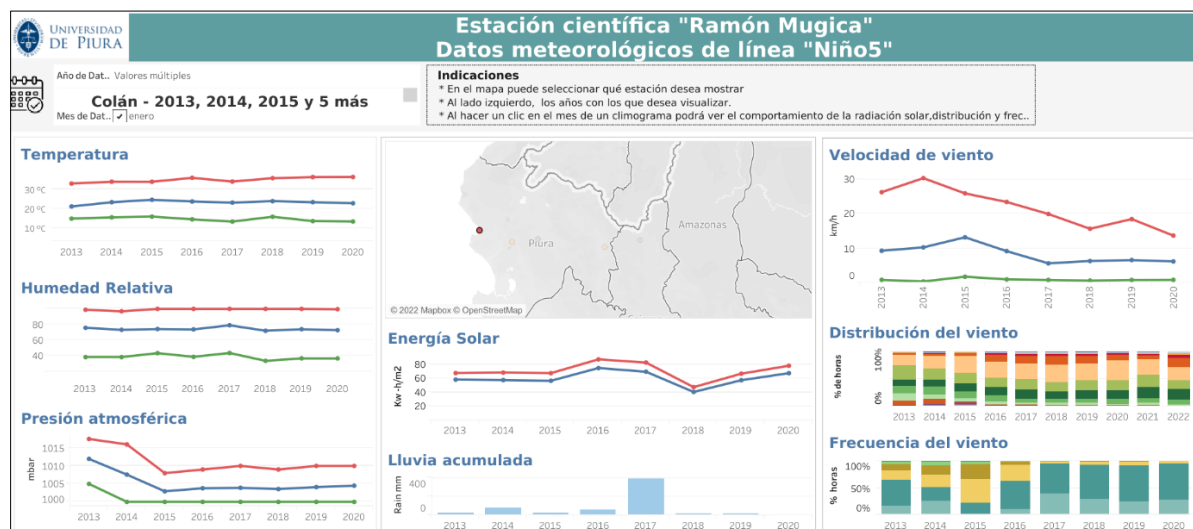
Figura 49. Dashboard mensual Niño 5



² Tableau public: https://public.tableau.com/views/DashboardNio05/Dashboardclima?:language=es-ES&display_count=n&origin=viz_share_link

³ Tableau public: https://public.tableau.com/views/DashboardNio5_year/Dashboardclima?:language=es-ES&display_count=n&origin=viz_share_link

Figura 50. Dashboard anual Niño5



Es importante recalcar que los flujos de integración, limpieza y ordenamiento están diseñados para realizar su ejecución de manera automatizada cuando se ingresa una nueva fuente de datos. De igual manera con las visualizaciones y Dashboards, estas se actualizan automáticamente, pero siempre y cuando la fuente de datos inicial recorra todos los flujos en el orden correcto.

3.4 Receptores de información

Una vez terminados los procedimientos de obtención, preparación y visualización de la información, se procede a decidir la relación de *stakeholders* y a qué información tendrán acceso según el rubro al que pertenecen. La importancia de esto radica en las preguntas de qué compartir y a quién compartir. Dado que, si se entrega información de parámetros innecesarios a los interesados sería contraproducente.

Es importante destacar que previo a la elección de *stakeholders*, se elaboraron las visualizaciones de acuerdo con las necesidades estándar sugeridas y aprobadas por el director de la EC, Rodolfo Rodríguez quien ha tenido un involucramiento activo en este trabajo. Asimismo, la CSI permitirá la elaboración de nuevas visualizaciones acorde con las nuevas necesidades de los *stakeholders* identificados.

La información a compartir la categorizaremos en 3 partes: (1) Fuentes de datos; (2) Flujos de preparación (3) Visualizaciones. En los flujos de preparación estarán los procedimientos establecidos en integración, limpieza y ordenamiento.

En principio el criterio de elección de *stakeholders* ha sido definido por el director de EC Ramón Mugica, a los cuales hace referencia con instituciones con las que se trabajado en etapas anteriores. Como el IGP, SENAEMI. Asimismo, uno de los fines del trabajo en la EC es

aportar al desarrollo de la investigación regional por lo que es necesario definir a interesados para el trabajo complementario con otras universidades y tesis. (Ver Tabla 14)

Tabla 14. Relación de *stakeholders*

N°	Stakeholders	Información		
		Fuente de datos	Flujos de preparación	Visualizaciones
1	Director EC Ramón Mugica	X	X	X
2	Equipo de Investigación EC Ramón Mugica	X	X	X
3	Tesistas pregrado, posgrado y doctoral (UDEP)	X		X
4	Instituciones IGP, SENAHMI, Ministerio de Salud, Gobierno Regional, COER Piura, Municipalidad Provincial Piura, Ministerio de agricultura, Autoridad Nacional del Agua, investigadores UNP.	X		X
5	Público general			X

Esta etapa de distribución se implementará con Tableau Online, el cual permitirá la creación de usuarios con distintos roles y acceso a los diferentes flujos y libros de trabajo, en los cuales se encuentran las visualizaciones de Tableau, desarrollados en esta cadena de suministro de información. Los flujos de preparación contendrán los flujos de integración, limpieza y ordenamiento. De igual manera, los *dashboards* se compartirán de manera pública en la plataforma de Tableau Public con la finalidad de que el público en general pueda visualizar el comportamiento de los parámetros meteorológicos estudiados para cada estación seleccionada sin alguna restricción, pero sin la opción de descargar los datos. Para acceder a estos solo se les comparte un enlace de Tableau Public sin la necesidad de ser un usuario con licencia. (Ver nota de pie 2 y 3).

Cada usuario creado tendrá un rol asignado en Tableau Online, lo que le permitirá tener ciertos permisos sobre la información compartida como el de editar y publicar nuevos flujos de preparación y/o visualizaciones, así como compartir estos flujos por medio de la plataforma de Tableau Online con otros usuarios. Igualmente, la opción de descargar estos flujos de preparación y/o visualizaciones. Es importante establecer una jerarquía con el equipo de trabajo, dado que en una cadena de suministro de información existe una serie de agentes

con tareas específicas en el manejo de información, en este caso son los integrantes adscritos al equipo de la EC Ramón Mugica.

Si bien es cierto, el rol de interesados en el tema meteorológico es muy amplio por la variedad de aplicaciones de estudio que ofrece, sin embargo, se integrarán como usuarios de Tableau Online a un grupo reducido de interesados potenciales como tesisistas, investigadores y/o personas relacionadas a las organizaciones mencionadas en la Tabla 14. Asimismo, se recalca la posibilidad de integración de nuevos usuarios según el permiso que atribuya el director de la EC.

La Tabla 15 muestra las actividades que puede hacer el usuario en los flujos y libros de trabajo de acuerdo con el rol asignado. El director de la EC y su equipo de investigadores tendrán roles como creadores y exploradores. El resto trabajará como visualizadores de datos y gráficos.

Tabla 15. Roles de usuarios registrados

Actividad/Rol	Creador	Explorador	Visualizador
Visualizar	X	X	X
Interactuar	X	X	X
Descargar imágenes/PDF	X	X	X
Descargar datos resumidos	X	X	X
Editar flujos/visualizaciones	X	X	
Publicar flujos/visualizaciones	X	X	
Mover flujos/visualizaciones	X	X	
Eliminar flujos/visualizaciones	X	X	
Configurar permiso	X		

Es importante añadir que, si un usuario desea acceder a otra actividad, pero su rol no lo permite se le puede atribuir un rol con una configuración personalizada en donde pueda habilitarse una actividad extra a su rol, siempre y cuando el usuario creador se lo asigne.

La distribución de la información aplicada en base a la asignación de usuarios y roles facilitará los procesos de manejo de información, debido a que los agentes asignados como creadores y exploradores podrán manejar los procedimientos de manera complementaria e interdependiente de acuerdo con los requerimientos de información que se plantean obtener previamente.

3.5 Implementación

Previamente a la asignación de usuarios se creará un proyecto en Tableau Online denominado “Red Niño 5”, en donde los usuarios con rol creador podrán subir todos los archivos flujos y libros de trabajo realizados en el presente trabajo.

Los archivos subidos en la plataforma se categorizarán por base de datos, flujos de integración, flujos de limpieza, flujos de ordenamiento y libros de trabajo.

Los flujos de integración, limpieza y ordenamiento se suben desde Tableau Prep Builder, en donde utilizando la opción de “Publicar flujo” se tendrá acceso a Tableau online, siempre y cuando el usuario sea un creador o explorador con opción a subir flujos de trabajos. (Ver Figura 51)

Figura 51. Pasos para subir flujos a Tableau Online

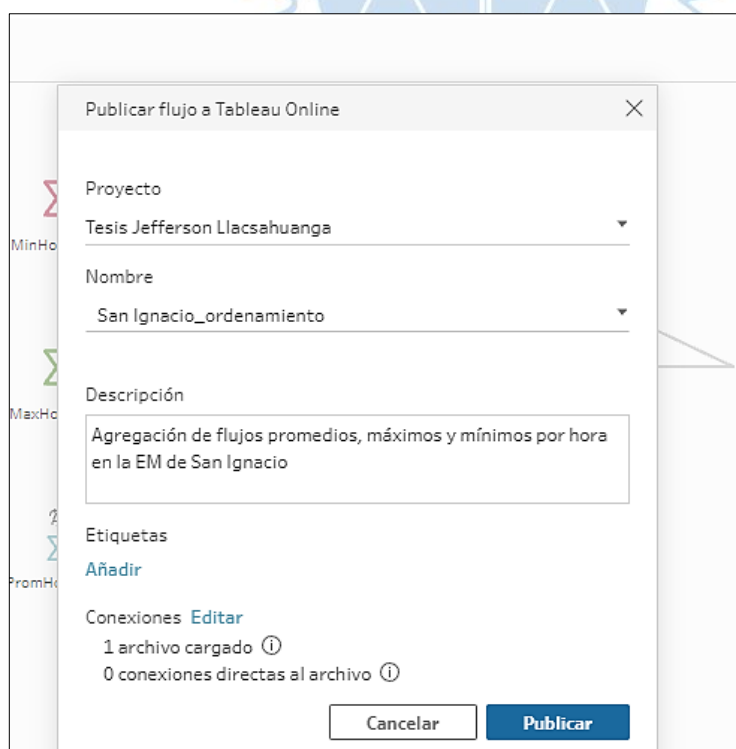


Figura 52. Dashboard Tableau Online



Reglas de permisos									
Grupo/usuario	Plantilla								
Alexander Vasquez ***	Explorar ▼								
Carla Piedra ***	Publicar ▼								
Carlos Fabián Torres Guerra ***	Ver ▼								
Mario Quinde ***	Administrar ▼								

+ Añadir grupo/Regla de usuario

Presentación de resultados

Como parte de un criterio de validación se estableció presentar las visualizaciones a un público con la finalidad de interactuar y obtener *feedbacks* respecto a la presentación de la información. Esto se realizó en 6 oportunidades tal como figura en la Tabla 16, los cuales fueron de gran importancia, ya que por sus opiniones respectivas se implementaron las mejoras en los filtros de años y meses para tener una mayor interacción con la información.

Tabla 16. Presentación de visualizaciones

Fecha	Público
9 de junio 2022	Visita técnica de universitarios de la facultad de Ingeniería UDEP
24 de junio 2022	Visita encargados de Municipalidad Provincial de Piura.
8 de julio 2022	Tesista doctoral de Universidad Nacional de Piura
19 de agosto 2022	Visita técnica de universitario de la facultad de Física de la Universidad Nacional de Piura.
29 de octubre 2022	Visita de Maestría de Ingeniería ambiental y seguridad Industrial de la UNP
4 de noviembre 2022	Visita de delegación holandesa como parte del evento del <i>Climate Café</i> junto a estudiantes de la Universidad de Piura, Universidad Cesar Vallejo y Universidad Nacional de Piura.

Conclusiones

La importancia de la CSI no solo radica en la implementación de nuevas metodologías de explotación de datos. También en procesos enfocados al ciclo de vida de los datos, dado que en el rubro de investigación meteorológica se trabaja con un gran volumen y variedad de estos, ya que va de la mano con los rangos de años, cantidad de parámetros correlacionados, y número de estaciones meteorológicas a estudiar. La CSI definirá los pasos específicos para el proceso de preparación, visualización y distribución de datos de acuerdo con los distintos requerimientos de información solicitados, evitando así los métodos tradicionales de manejo de datos.

La CSI implementada mejora la eficiencia en tiempo y recursos desde la preparación hasta la distribución de datos, ya que los agentes encargados del manejo de información pueden optar por aplicar procedimientos personalizados y variados con referencia a los requerimientos demandados por el razonamiento que se desea emplear para obtener las aplicaciones planteadas inicialmente como el análisis del comportamiento de parámetros como la temperatura, lluvia, entre otros, en un determinado tiempo. Además, este análisis se puede realizar de manera paralela y complementaria según las metodologías adquiridas por el equipo de trabajo.

La implementación de la CSI logra la reducción de horas de trabajo empleadas en la preparación de datos, ya que antes se destinaban un aproximado de más de 8 horas utilizando los métodos tradicionales con las hojas de trabajo de Microsoft Excel y lenguajes de programación, sin embargo, con esta implementación se trabaja en menos de una hora. Facilitando los análisis descriptivos y predictivos posteriores.

La eficiencia del análisis de los datos meteorológicos dependerá del correcto procedimiento de preparación de datos establecido en la cadena de suministro de información. Dado que, si se emplea procedimientos errados en la limpieza de datos sin tener en cuenta datos atípicos o correctos criterios de reemplazo de datos en parámetros específicos, se obtendrán resultados totalmente ajenos y confusos. Esta tendría como consecuencia una alteración en los requerimientos de información solicitados e incluso una mala toma de decisiones de acuerdo con la finalidad de la obtención de la información.

El proceso de preparación fue el que tomó una mayor relevancia en el desarrollo de este trabajo, debido a que, al iniciar con la integración de datos, estos se presentaban en formatos variados. Además, en la limpieza de datos se debe definir de manera detallada las características que deben cumplir para ser considerados errados; de lo contrario, si esto se hace de manera incorrecta, afectará en el análisis y visualizaciones posteriores.

La implementación de la CSI en la estación científica Ramón Mugica es modulable y escalable, lo que la hace una propuesta de alto impacto en la preparación de datos. Los procedimientos de preparación establecidos inicialmente pueden llegar a ser modificados de acuerdo con el análisis que se busca obtener. Es importante añadir que esta implementación cumple con el propósito de optimizar esfuerzos en el equipo de investigación de la estación científica para la preparación de datos, ya que los flujos implementados son adaptables a los requisitos que el usuario busca obtener.

Los requerimientos de información sobre datos meteorológicos deben estar establecidos previamente a distribuir la información, ya que, en el rubro meteorológico existen una variedad de aplicaciones de estudio que se pueden realizar, siempre y cuando se emplee los razonamientos correctos junto a la preparación de datos adecuada. En suma, en la distribución de información se irán encontrando una serie de interesados potenciales a lo largo del tiempo, por lo que la CSI facilitará el acceso de una manera escalable a próximas organizaciones y/o usuarios individuales que deseen interactuar con esta información para distintos fines, en cuanto se verifique un fin complementario a la investigación o útil a la sociedad.

Se resalta la importancia del monitoreo de los datos meteorológicos para la preparación ante la ocurrencia de eventos climáticos, por lo que una CSI bien establecida con los procedimientos correctos facilitará aumentar el estudio de otros parámetros, así como el de otras estaciones meteorológicas. Especialmente, las que se encuentran ubicadas en Paita y registran la temperatura del agua del mar. Esta implementación aportará a realizar un análisis comparativo y correlacional para aportar en la toma de decisiones preventivas ante los fenómenos de El Niño y La Niña, siendo estos los más comunes en el norte del Perú.

En la limpieza de datos se pueden añadir nuevas metodologías para reemplazar datos nulos, como el reemplazo de datos obtenidos de estaciones meteorológicas cercanas. Si se requiere hacer un análisis comparativo y descriptivo más a detalle, se necesitará tener una data mucho más íntegra para obtener conclusiones adecuadas sobre el comportamiento de parámetros específicos. Esto aplica especialmente para los en donde las estaciones meteorológicas se averiaron por un evento climático.

La estación científica Ramón Mugica no solo posee un gran volumen de fuente de datos de las estaciones meteorológicas, sino también de otros proyectos anexados a ella. Por tal motivo, la implementación de la CSI también se puede aplicar a nuevos trabajos de investigación utilizando el mismo software y procedimiento, siempre y cuando se modifiquen algunos criterios relacionados al comportamiento de los nuevos datos a trabajar.

Se sugiere una obtención de datos meteorológicos de manera remota en todas las estaciones meteorológicas, ya que facilitaría la actualización de la fuente de datos en un rango más breve de tiempo. De igual manera, como trabajo futuro, se debe mejorar continuamente la cadena de suministro de información para poder monitorear los datos en tiempo real y, así, realizar análisis descriptivo y predictivo que soporten la toma de decisiones en acciones de prevención climática, potencial energético eólico y solar, salud, entre otros.





Glosario

BA: Business Analytics

JIT: Just in time

FD: Fuente de datos

CSI: Cadena de suministro de Información

EC: Estación científica

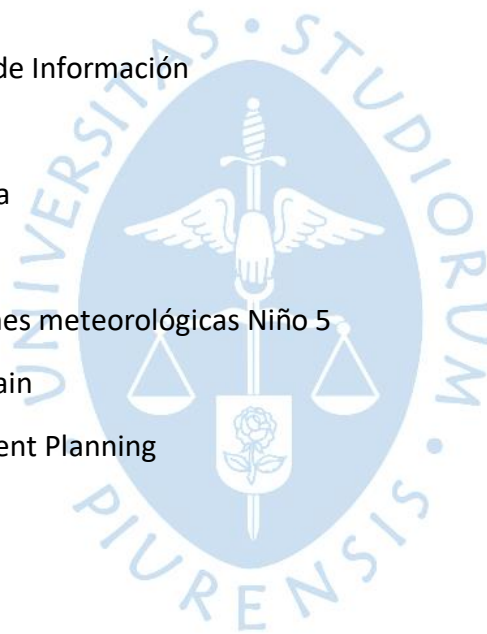
EM: Estación meteorológica

FEN: Fenómeno El Niño

NIÑO5: Cadena de estaciones meteorológicas Niño 5

ISC: Information Supply Chain

IRP: Information Requirement Planning





Referencias bibliográficas

- Davenport , T. (2007). *Competing on Analytics*.
- Davenport, T. (Diciembre de 2013). *Analytic 3.0. Harvard Business Review*. Recuperado el 14 de Julio de 2022
- Hijar, G., Bonilla, C., Munayco, C., Gutierrez, E., & Ramos, W. (2016). *Fenómeno El Niño y Desastres Naturales: Intervenciones en salud pública para la preparación y respuesta*. Rev Peru Med Exp Salud Publica. doi:10.17843/rpmpesp.2016.332.2205
- Hindle, G., Kunc, M., Mortensen, M., Oztekin, A., & Vidgen, R. (2019). *Business analytics: Defining the field and identifying a research*. doi:10.1016/j.ejor.2019.10.001
- Holsapple, C., Lee-Post, A., & Pakth, R. (2014). *A unified foundation for business analytics*. doi:10.1016/j.dss.2014.05.013
- Instituto del Mar del Perú (IMARPE). (2018). *El Niño*. Recuperado el 13 de Agosto de 2022, de
nstituto del Mar del Perú:
http://www.imarpe.gob.pe/imarpe/index2.php?id_seccion=I017801000000000000000
- Kovats, R., Bouma, M., Hajat, S., Worall, E., & Haines, A. (2003). *El Niño and health*. doi:10.1016/S0140-6736(03)14695-8
- Nagle, T., & Sammon, D. (07 de Marzo de 2016). *Business & Finance*. Obtenido de Business & Finance: <https://businessandfinance.com/blog/guest-blog-being-data-savvy/>
- Omar , Y., Minoufekar, M., & Plapper, P. (2019). *Business analytics in manufacturing: Current trends, challenges and pathway to market leadership*. doi:10.1016/j.orp.2019.100127
- Perera , C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2014). *Context Aware Computing for The Internet of Things: A Survey*. IEEE Communications Surveys & Tutorials. doi:10.1109/SURV.2013.042313.00197
- Rodriguez , R., & Marcelo, D. (2009). *Seis estaciones meteorológicas para el estudio de los efectos del fenómeno climático El Niño en las actividades productivas y en la población*. Proyecto, FINCYT. doi:05-2008/FINCYT/EQ

- Rodriguez, R. (2011). *El Fenómeno El Niño- Oscilación sur*. Publicación, Universidad de Piura. Recuperado el 13 de Agosto de 2022
- Rodriguez, R. (14 de Marzo de 2013). *Docentes de la Facultad de Ingeniería participaron de esta reunión científica con la exposición de proyectos de investigación agraria que generan impacto en la Región*. (O. O. Córdova, Editor) Recuperado el 14 de Agosto de 2022, de Udep: <https://www.udep.edu.pe/hoy/2013/03/expertos-de-ingenieria-presentaron-proyectos-ejecutados-en-piura/>
- Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI). (2014). *Fenómeno EL Niño en el Perú*. Recuperado el 13 de Agosto de 2022, de https://www.minam.gob.pe/wp-content/uploads/2014/07/Dossier-El-Ni%C3%B1o-Final_web.pdf
- Sun, S., & Yen, J. (2005). *Information Supply Chain: A Unified Framework for Information-Sharing*. doi:10.1007/11427995_38
- Tableau. (2020). *Tableau Learn*. Recuperado el 8 de Agosto de 2022, de Tableau: <https://www.tableau.com/learn/articles/what-is-data-cleaning>
- Takahashi, K. (2014). *Variedades de El Niño*. Boletín Técnico, Generación de modelos climáticos para el pronóstico de la ocurrencia del Fenómeno El Niño, Instituto Geofísico del Perú. Recuperado el 13 de Agosto de 2022, de <http://hdl.handle.net/20.500.12816/5042>
- Takahashi, K. (2017). *Fenómeno El Niño: "Global" vs "Costero"*. Boletín técnico: Generación de información y monitoreo del Fenómeno El Niño, Insituto Geofísico del Perú. Recuperado el 13 de Agosto de 2022, de <http://hdl.handle.net/20.500.12816/5101>
- Zambrano, E. (1986). *El fenómeno de "El Niño" y la Oscilación Sur (ENSO)*. Acta Oceanográfica del INOCAR, Guayaquil. Recuperado el 13 de Agosto de 2022