

Bibliotecas de
Python



Herramientas de
Visualización de datos

Sprint 2

“Herramientas de estadísticas y
probabilidad”

Gerardo Javier Ortega
Iturbe

09

10

2024

Data Analyst
Fundamentals



NAO ID: 3149



Backlog

Lista de Requerimientos

Historias de usuario	Requerimiento
<ul style="list-style-type: none"> Yo como Ingeniero de Datos requiero acceso a los datos para extraer la información. 	Requiero tener acceso a la base de datos, así como las credenciales necesarias.
<ul style="list-style-type: none"> Yo como Analista de Datos requiero el archivo de datos para realizar transformaciones. 	Se solicita el archivo en formato csv.
<ul style="list-style-type: none"> Yo como Analista de Datos requiero realizar gráficos para identificar los factores asociados al problema 	Se requieren la tabla consolidada
<ul style="list-style-type: none"> Yo como Analista de datos requiero los datos para realizar el análisis exploratorio. 	Se debe contar con los datos limpios listos para el análisis.
<ul style="list-style-type: none"> Yo como Analista de Datos requiero el script para almacenar los datos. 	Se requiere el script para almacenar los datos
<ul style="list-style-type: none"> Yo como Analista de Datos requiero los datos para realizar las uniones. 	Se requieren el acceso a las tablas
<ul style="list-style-type: none"> Yo como Analista de Datos requiero crear un programa que genere histogramas comparativos de las ventas de órdenes completas que tuvieron retrasos moderados y prolongados. 	Se requieren la tabla consolidada
<ul style="list-style-type: none"> Como analista de datos, requiero escribir un programa que cree un mapa interactivo que muestre el valor monetario de los retrasos prolongados por estado 	Se requieren la tabla consolidada
<ul style="list-style-type: none"> Como analista de datos, requiero desarrollar un programa que construya una matriz de correlación entre la variable delta_days y otras variables numéricas, para analizar las relaciones entre los retrasos y otros factores en órdenes completas. 	Se requieren la tabla consolidada
<ul style="list-style-type: none"> Como analista de datos, quiero crear un programa interactivo que muestre la evolución mensual y anual de las órdenes con retrasos prolongado. 	Se requieren la tabla consolidada

<ul style="list-style-type: none">● Como analista de datos, requiero acompañar cada visualización con una interpretación escrita que incorpore los insights obtenidos.	Se requieren las visualizaciones obtenidas en los scripts
<ul style="list-style-type: none">● Como analista de datos, necesito un programa que genere una visualización de la distribución de delta_days en los estados de Brasil, segmentada por el estado del retraso (delay_status).	Se requieren la tabla consolidada
<ul style="list-style-type: none">● Como analista de datos, quiero desarrollar un programa que visualice un mapa interactivo que represente la cantidad de órdenes con retrasos prolongados por estado en Brasil	Se requieren la tabla consolidada

Lista Priorizada

Requerimientos	Etapas	Estimación de Tiempo	Entregable
<input checked="" type="checkbox"/> Yo como Ingeniero de Datos requiero acceso a los datos para extraer la información.	Sprint 1	1 hora	Script
<input checked="" type="checkbox"/> Yo como Analista de Datos requiero el archivo con los datos para ejecutar la limpieza de datos.	Sprint 1	2 horas	Script
<input checked="" type="checkbox"/> Yo como Analista de Datos requiero el archivo de datos para realizar transformaciones.	Sprint 1	2 horas	Script
<input checked="" type="checkbox"/> Yo como Analista de datos requiero los datos para realizar el análisis exploratorio.	Sprint 1	2 horas	Script
<input checked="" type="checkbox"/> Yo como Analista de Datos requiero los datos para realizar las uniones.	Sprint 1	2 horas	Script
<input checked="" type="checkbox"/> Yo como Analista de Datos el script para almacenar los datos.	Sprint 1	1 hora	Código fuente
<input checked="" type="checkbox"/> Yo como Analista de Datos requiero realizar gráficos para identificar los factores asociados al problema	Sprint 1	8 horas	Documento con interpretaciones, gráficos y script.
<input type="checkbox"/> Yo como Analista de Datos requiero crear un programa que genere histogramas comparativos de las ventas de órdenes completas que tuvieron retrasos moderados y prolongados.	Sprint 2	2 horas	Imagen estática llamada "3_a_histogram_sales_short_long_delays.png"
<input type="checkbox"/> Como analista de datos, requiero desarrollar un	Sprint 2	2 horas	Imagen estática llamada "3_b_correlation_matrix_

<p>programa que construya una matriz de correlación entre la variable delta_days y otras variables numéricas, para analizar las relaciones entre los retrasos y otros factores en órdenes completas.</p>			complete_orders.png”
<input type="checkbox"/> Como analista de datos, necesito un programa que genere una visualización de la distribución de delta_days en los estados de Brasil, segmentada por el estado del retraso (delay_status).	Sprint 2	2 horas	Imagen estática llamada “3_c_delta_day_by_state_and_delay_type.png”
<input type="checkbox"/> Como analista de datos, quiero crear un programa interactivo que muestre la evolución mensual y anual de las órdenes con retrasos prolongados.	Sprint 2	2 horas	Imagen llamada "3_d_evolution_delayed_orders_by_region.html”
<input type="checkbox"/> Como analista de datos, quiero desarrollar un programa que visualice un mapa interactivo que represente la cantidad de órdenes con retrasos prolongados por estado en Brasil	Sprint 2	2 horas	Imagen llamada "3_e_map_long_delays_by_state.html"
<input type="checkbox"/> Como analista de datos, requiero escribir un programa que cree un mapa interactivo que muestre el valor	Sprint 2	2 horas	Imagen llamada "3_f_map_monetary_value_long_delays_by_state.html"

monetario de los retrasos prolongados por			
<input type="checkbox"/> Como analista de datos, requiero acompañar cada visualización con una interpretación escrita que incorpore los insights obtenidos.	Sprint 2	4 horas	Documento PDF con interpretaciones titulado "2_interpretation.pdf"



Road Map

Sprint 2 “Herramientas de estadísticas y probabilidad”

Objetivos Generales

- Extraer y preparar los datos.
- Limpieza y transformación de la información.
- Obtención de insights
- Creación de gráficos

Objetivos Específicos

- Llevar a cabo la limpieza de los datos para eliminar la redundancia y mantener la calidad.
- Implementar un proceso de transformación de los datos para asegurar su correcto manejo.
- Almacenar los datos procesados para facilitar su posterior consulta y análisis.
- Desarrollo de gráficos por Estado y año
- Desarrollo de gráficos interactivos de ventas con retrasos prolongados
- Generación de un documento con los insights

Proceso	Sprint 1 02/10/2024	Sprint 2 09/10/2024	Sprint 3 15/10/2024
Extracción de Datos	Poner a disposición del equipo de análisis de datos los datos.		
Limpieza de Datos	Eliminar duplicados, manejo de valores nulos.		
Transformación de Datos	Creación/ eliminación de columnas, normalización de texto, conversión a formato correcto.		
Almacenamiento	Entrega del código fuente y archivo generado.		
Desarrollo de gráficos		Desarrollo de gráficos que reflejen la problemática y su interpretación	
Reporte		Se resumirán los principales insights obtenidos a partir del análisis.	