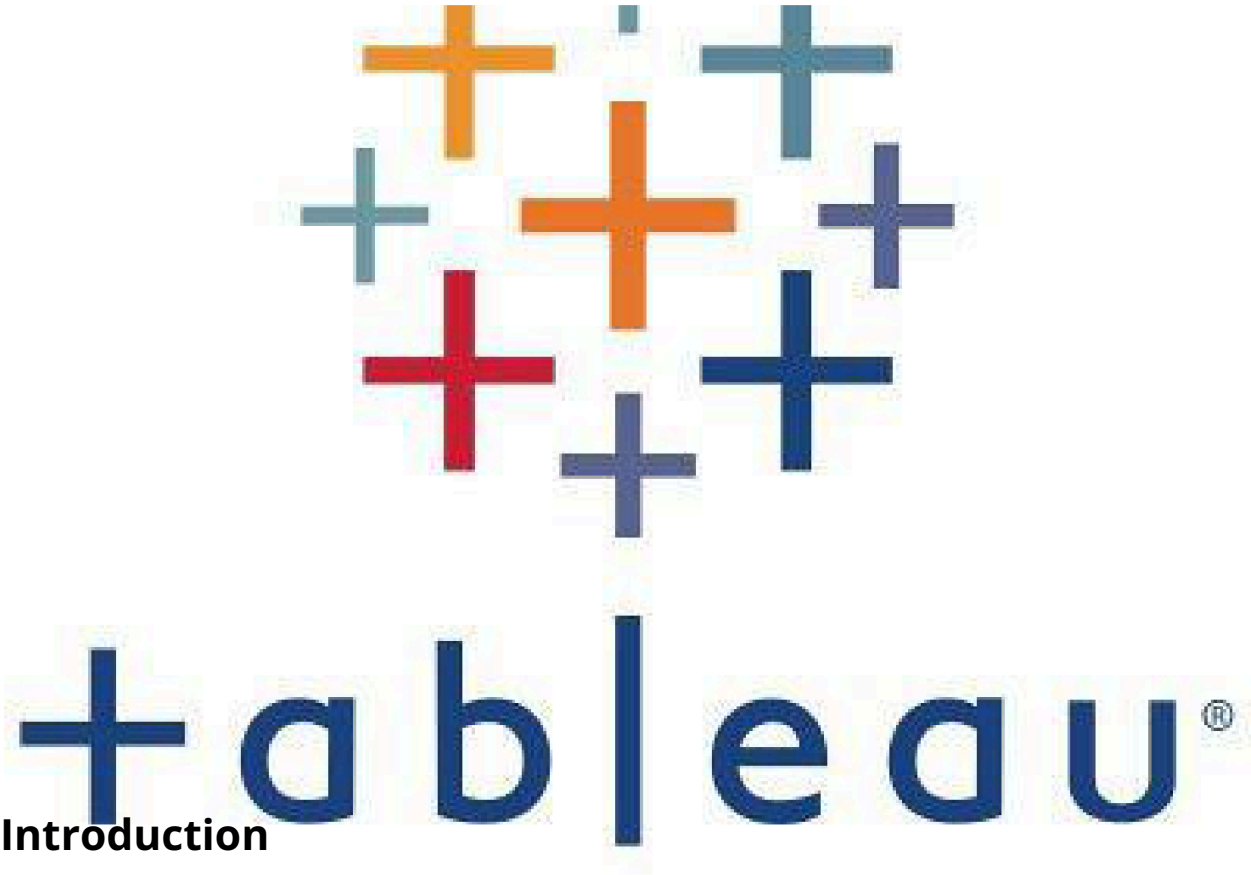
Data visualization

REPORT

Exploratory Data Analysis in Tableau



Introduction

This report provides an in-depth analysis of rainfall data in Australia(dataset1) and influenza dataset(dataset 2), focusing on various meteorological parameters and their

relationship with temperature. The dataset 1 encompasses a wide range of information collected at different times of the day, offering valuable insights into weather patterns and trends.

The dataset 2 includes information on the number of influenza cases, subtypes, and related attributes. The primary goal of this analysis is to gain insights into influenza trends, subtype distribution, and their potential correlation with external factors such as temperature anomalies. The datasets used for these analyses have been meticulously cleaned and prepared for analysis.

Data cleaning

In these datasets, there were many null values in some columns. Not all null values should be removed without consideration. Removing rows with null values can lead to the loss of potentially important data in other columns.

During the cleaning process, I evaluated the relevance of each column for our analysis objectives. Columns that did not contribute to our specific research questions were removed.

Prior to analysis, the dataset underwent rigorous data cleaning procedures to ensure data accuracy and consistency. This involved correcting anomalies, and formatting data for visualization.

The column names in the Rain in Australia dataset were user friendly and easy to understand but for the influenza dataset, it was the complete opposite. I changed the column names for easy understanding of the user.

SPEC_RECEIVED_ND: "Unsubtyped Influenza Specimens Received"

SPEC_PROCESSED_NB: "Influenza Specimens Processed but Not Subtyped"

AH1: "Influenza A (H1N1) Positive Cases/Specimens"

AH1N12009: "Influenza A (H1N1pdm09) Positive Cases/Specimens (2009 Strain)"

AH3: "Influenza A (H3N2) Positive Cases/Specimens"

AH5: "Influenza A (H5N1) Positive Cases/Specimens (Avian Influenza)"

ANOTSUBTYPED: "Influenza A Cases/Specimens Not Subtyped"

INF_A: "Total Influenza A Positive Cases/Specimens"

BYAMAGATA: "Influenza B (Yamagata Lineage) Positive Cases/Specimens"

BVICTORIA: "Influenza B (Victoria Lineage) Positive Cases/Specimens"

BNOTDETERMINED: "Influenza B Cases/Specimens Not Further Classified"

INF_B: "Total Influenza B Positive Cases/Specimens"

ALL_INF: "Total Influenza A and B Positive Cases/Specimens"

ALL_INF2: "Additional Data on All Influenza Cases/Specimens"

Insights:

Certainly, here are some questions answered through analysis :

Influenza dataset:

- Are there specific years or seasons where a particular subtype, such as H1N1 or H3N2, was more dominant?
- How has the total number of influenza cases, including both influenza A and B, changed over time?
- Are there any noticeable patterns or fluctuations in the total cases?
- What is the proportion of influenza cases that could not be subtyped ("Unsubtyped Influenza Specimens Received" and "Influenza Specimens Processed but Not Subtyped")?
- Are there significant regional variations in the prevalence of influenza cases?● Can we identify regions with consistently higher or lower influenza activity?

3

- How does the distribution of influenza B cases compare between the Yamagata and Victoria lineages?
- Is there any correlation between temperature anomalies and the number of influenza cases?
- Do extreme temperature conditions coincide with increased influenza activity?
- Are there specific years with exceptionally high influenza activity?
- Does the impact of temperature anomalies on influenza cases vary by region?
- Do certain subtypes exhibit stronger associations with specific temperature conditions?

Rain in Australia:

- How does wind speed vary between morning (9 am) and afternoon (3 pm)?
- Are there any instances of extreme wind speeds that coincide with specific temperature ranges?
- Are there temperature ranges associated with high or low rain pressure?
- Is there a correlation between humidity levels and temperature?
- Are there specific times of the day when humidity tends to be higher or lower, and how does it relate to temperature?
- How do morning (9 am) and afternoon (3 pm) temperatures compare on average?
- How do daily minimum and maximum temperatures correlate with average temperatures?
- Which city in Australia experiences the highest total rainfall, and during which months?
- How does the duration of sunshine relate to temperature variations?
- Are there temperature ranges associated with longer or shorter periods of sunshine?

Conclusion:

In conclusion, Tableau was used to preprocess and analyze data related to rain in Australia and influenza regions. The data cleaning process involved various tasks mentioned above using Tableau Prep. The datasets were then blended in Tableau Desktop to draw insights from the data. The visualizations and answers to user questions provide a better understanding of weather patterns and trends, contributing to informed decision-making in various fields for dataset 1 and for dataset 2 it provided valuable insights into the distribution of influenza subtypes, the burden of influenza A and B cases, and regional variations in influenza prevalence. Both the datasets' correlation with temperature is shown in stories.

DATA SOURCE:

y