BIKE RIDE DATA PORTFOLIO IN LONDON DATASET

TABLE OF CONTENTS

1. Introduction

Purpose of Bike Riding Data Portfolio in London Dataset:

The introduction section acts as a point of entry, providing a thorough knowledge of the core goals of the London Bike Rides Dashboard. It clarifies how the bike-sharing dashboard is essential to comprehending the dataset that is offered.

Scope and Objectives:

Through this interactive dashboard, the way to users's understanding of bike sharing taking place in London will become a lot easier. The user can easily identify when most bike rides take place, the trends and the pattern according to the weather and the seasonal changes. The objectives of this analysis are outlined, shedding light on the overarching goals, such as enhancing user experience and improving the visual perspective of the dataset. By clearly defining the scope and objectives, the report sets the stage for a focused and meaningful exploration of the data.

Data Sources:

Within this section, an intricate overview of the data sources employed for the creation of the bike ride dashboard is provided. This encompasses a detailed description of the types of data procured, ranging from transactional records and customer demographics to market trends and product inventory. Emphasis is placed on the relevance of these data sources to the bike-sharing analysis, elucidating how they collectively contribute to a holistic understanding of user engagement in bike-sharing. Through this comprehensive explanation, readers gain insight into the robust foundation upon which the subsequent analyses and visualizations are constructed, ensuring transparency and credibility in the findings presented.

2. Data Analysis and Preparation

Data Collection and Sources:

The data underpinning this Bikeride dashboard was meticulously gathered from a myriad of sources and meticulously curated to ensure its accuracy and relevance. A combination of visuals, capturing various patterns among bike riders, and external circumstances such as temperature humidity windspeed etc are compared with each other to bring out the relation between them. This amalgamation of internal and external data sources enriched the dataset, providing a panoramic view of the field dynamics and customer behaviour crucial for nuanced analysis.

Data Cleaning and Preprocessing:

Prior to delving into the analytical processes, the raw data underwent a rigorous cleaning and preprocessing regimen. Robust data cleaning algorithms were employed to detect and rectify discrepancies, ensuring data integrity. Missing values were meticulously handled through imputation techniques, ensuring that no gaps hindered the analysis. Duplicates were systematically removed, guaranteeing a singular representation of each data point. The data was then transformed into a consistent format, harmonizing disparate data types and units. Additionally, outliers were identified and either corrected or removed, preventing skewed interpretations. This meticulous preprocessing not only fortified the dataset’s reliability but also streamlined subsequent analytical endeavors.

Data Exploration and Descriptive Statistics:

The exploratory data analysis (EDA) phase unearthed compelling insights from the prepared dataset.Visualizations, including line charts,heatmaps and bar charts were crafted to elucidate data distributions, revealing patterns and anomalies. Initial insights gleaned from the dataset shed light on seasonal or weather preferences, daily trends, and circumstantial effects. For instance, a notable surge in bike rides during specific seasons became apparent, correlating with weather, season and other external factors. In essence, this meticulous data analysis and preparation process not only ensured the dataset's reliability and coherence but also laid the foundation for the subsequent sections of this report. The insights gleaned from this phase serve as the bedrock upon which strategic decisions and actionable recommendations will be formulated, propelling the organization towards its sales objectives with newfound clarity and confidence

3. Dashboard Overview

Dashboard Layout and Structure:

The bike ride dashboard is meticulously crafted with a user-centric design, ensuring an intuitive and insightful user experience. The layout is organized into distinct sections, each catering to a specific aspect of the bike-sharing analysis. There will be a total of 5 visualizations on the dashboard, consisting of 3 primary visualizations and 2 additional visualizations in the tooltip. The users will be able to view the total number of bike rides between the tenure, starting from the top left. A moving average chart which is a line chart is displayed in the top right corner of the count. Below them is a heat map showing the relationship between temperature and wind speed.

The total number of bike rides broken down by the hour and the weather may be seen when the user hovers over the moving average chart or the heat map.

A collection of actions and filters, including user-defined parameters, will be included in the main visual, the moving average chart. The number is the first of these parameters, while the second one is the period—days, weeks, or months, depending on the user's preference. The timeline can be filtered with a third parameter, allowing you to shorten or lengthen it based on your preferences.The dashboard aims to facilitate a comprehensive analysis of London bike ride data, allowing users to uncover trends, correlations, and patterns in the dataset.

Interactive Features:

Holistic Analysis: Give customers a comprehensive understanding of London's bike travel patterns by considering variables including the time of day, the weather, and general trends.

Trend Identification: By reducing short-term oscillations, the Moving Average Rides Graph helps users discern long-term trends and gain a better understanding of overall patterns.

Date-specific Analysis: By enabling users to examine bike ride counts within time frames, the Total Rides Count visualization can help with the evaluation of anomalies

and short-term fluctuations.
Weather Impact: Using the Heat Map and Weather Bar Chart, show how the weather affects the number of bike rides, helping users to draw the dots between weather and ride patterns.

Time-of-day Insights: The Hour Time Bar Chart helps users plan ahead and discover when bike trips are most common by highlighting peak hours of bike traffic.These objectives.

Engaging Tooltips: Increase user satisfaction by adding interactive tooltips: Shifting Typical Rides Add the date and the moving rides average in the graph tooltip. Hovering on heatmap cells will reveal data about the temperature and wind speed, as well as information from the hour time bar chart and weather bar chart.

This comprehensive dashboard facilitates data-driven decision-making for a range of stakeholders, including city planners and bike-sharing businesses, by enabling users to obtain actionable insights into the behaviour of bike riders in London.

Technologies Used (Python Libraries):

The dashboard is developed using a robust stack of Python libraries and frameworks. Pandas, a powerful data manipulation library, is employed for data wrangling, cleaning, and transformation, ensuring the dataset is primed for analysis. A well-known open-source manipulation and analysis library. Along with capabilities for reading and writing several data formats, it offers data structures for effectively storing and processing massive datasets. The zipfile module in Python provides tools for working with ZIP archives. It allows you to create new ZIP files, extract files from existing ZIP archives, and more.

Data Visualization Techniques Employed:

The dashboard leverages a diverse array of data visualization techniques to cater to various analytical needs. The line chart is employed to depict the moving average on a daily/weekly/monthly basis, providing a clear view of trends followed during the selected period. Heat maps are utilized to visualize a relation between temperature vs windspeed, enabling users to weather and seasonal trends. Additionally, there are two visuals which cloud be accessed by hovering over the moving average line chart and the heatmap. These visualization techniques are chosen judiciously to address specific analytical questions, ensuring that the dashboard delivers comprehensive insights into the Bike ride analysis.

In summary, this dashboard stands as a testament to the synergy between advanced Python libraries, interactive features, and diverse visualization techniques. Its user friendly design, coupled with real-time data updates and interactive elements, empowers users to unravel intricate patterns within the bike ride data, driving informed decision-making and strategic planning within the organization.

4. Key Metrics and Visualizations

a) Moving Average Rides Graph:

Displays the rounded trend of bike trips over a period. This will assist in comprehending the entire image and may be useful in identifying any patterns. This is the primary visual and controls changes in other visuals as well.

b) Total Rides Count (Selected Dates):

Presents the total bike ride count. It will show the count within a user selected date range.

c) Heat Map (Temperature vs. Wind Speed):

Illustrates the correlation between temperature and wind speed, providing insights into potential weather influences on bike rides.

d) Weather Bar Chart (Color-coded):

Represents the distribution of weather conditions during bike rides, enhancing understanding of the impact of weather on ride counts.

e) Hour Time Bar Chart:

This shows the distribution of bike rides throughout the day, helping identify peak activity hours. User can use this to understand the underlaying hourly patterns and preferred timings by the bike riders.

5. In-Depth Analysis

Seasonal Ride Patterns: Delve deeper into seasonal patterns, identifying specific months or periods with significant bike ride spikes. Explore the reasons behind these patterns, such as holidays, promotions, or trends.

Performance Analysis: Conduct a detailed analysis of each Rider's performance, including when most bikers prefer to ride and in which weather. Identify peak hours and areas for improvement.

Rider Behavior Insights: Analyze bikers behaviour data, including preferred seasons and weather. Identify trends in customer preferences and behaviour in hourly analysis and temperature-windspeed relational visualization.

6. Challenges Faced

Data Quality Challenges:

Discuss challenges related to data quality, such as missing or inconsistent data. Describe the methods employed to address these challenges, such as data imputation or validation technique.

Technical Challenges:

During the development phase, various coding issues were identified, ranging from syntax errors to logical bugs within the algorithms. Rigorous code reviews and pair programming sessions were conducted to identify and rectify these issues promptly. Utilizing version control systems enabled collaborative debugging, ensuring that all team members were working with the latest bug-free codebase. Additionally, adopting a modular programming approach and adhering to best practices minimized the occurrence of coding issues.

User Experience Challenges:

The main issues with the user experience were with the dashboard's navigation and usability. challenges navigating the interface, locating features, and comprehending data visualizations. A thorough revamp of the dashboard's appearance and navigation process was carried out to address these issues. An interface that is easy to use and intuitive was created by implementing the principles of user-centered design. To assist users in navigating the dashboard's features, interactive tutorials, contextual tooltips, and clear and simple labeling were incorporated. Iterative usability testing sessions allowed for ongoing feedback to be included and the user experience improved.

7. Solution

A Python code is designed to manage and visualize the London Bike Rides dataset. It begins by leveraging the Kaggle library to access the dataset seamlessly through the Kaggle API. The dataset is downloaded in zip format and extracted using the zipfile library. The panda's library is then utilized to read the data from the Excel file, providing a convenient Data Frame for further processing.

The script incorporates a data cleaning and manipulation phase where specific adjustments can be made to enhance the dataset's quality and relevance. Following this, the class ids related to weather conditions and seasons are labeled. For instance, the script associates numeric weather codes with human-readable labels such as "Clear sky" or "Broken clouds."

To enhance the interpretability of the dataset, class ids for seasons, represented by numeric codes, are also labeled as "Spring," "Summer," "Autumn," or "Winter."

For visualization purposes, the script exports the cleaned and labeled data to a excel file named " **london_bikes_final.xlxs**" The script suggests using Tableau Public for creating visualizations based on this labeled dataset. Tableau Public is used for generating interactive and insightful visual representations of the data

8. Future Enhancements

I. Advanced Predictive Analytics:

Implement machine learning algorithms to forecast future ride trends based on historical data. Predictive models can provide valuable insights, enabling proactive decision-making and strategic planning.

II. Dynamic Drill-Down Capabilities:

Enhance interactivity by incorporating dynamic drill-down features. Allow users to explore data at a granular level, such as daily, weekly, or monthly ride trends. Implement intuitive navigation for seamless exploration.

III. Customizable Dashboards:

Empower users to customize their dashboards according to their specific needs. Implement drag-and-drop functionality, enabling users to add or remove widgets, rearrange visualizations, and save personalized dashboard layouts for future use.

IV. Geospatial Analytics:

Incorporate geospatial visualizations to analyze data in geographical contexts. Utilize maps to represent performance across regions, visualize locations, and identify high-bike-usage areas.

9. Code

Python code used to label the Dataset:

```
import pandas as pd
import zipfile
import kaggle

# download dataset from kaggle using the Kaggle API
 !kaggle datasets download -d
hmavrodiev/london-bike-sharing-dataset zipfile_name =
'london-bike-sharing-dataset.zip'
with zipfile.ZipFile(zipfile_name, 'r') as file:
    file.extractall()
bikes = pd.read_csv("london_merged.csv")

# specifying the column names that I want to use
new_cols_dict ={
 'timestamp':'time',
 'cnt':'count',
 't1':'temp_real_C',
 't2':'temp_feels_like_C',
 'hum':'humidity_percent',
 'wind_speed':'wind_speed_kph',
 'weather_code':'weather',
```

```
 'is_holiday':'is_holiday',
 'is_weekend':'is_weekend',
 'season':'season'
 }

# Renaming the columns to the specified column names
bikes.rename(new_cols_dict, axis=1, inplace=True)

# changing the humidity values to percentage (i.e. a value between 0 and 1)
bikes.humidity_percent = bikes.humidity_percent / 100

# creating a season dictionary so that we can map the integers 0-3 to the actual
written values
season_dict = {
 '0.0':'spring',
 '1.0':'summer',
 '2.0':'autumn',
 '3.0':'winter'
 }

# creating a weather dictionary so that we can map the integers to the actual
written values
weather_dict = {
 '1.0':'Clear',
 '2.0':'Scattered clouds',
 '3.0':'Broken clouds',
 '4.0':'Cloudy',
 '7.0':'Rain',
 '10.0':'Rain with thunderstorm',
 '26.0':'Snowfall'
 }

# changing the seasons column data type to string
bikes.season = bikes.season.astype('str')
# mapping the values 0-3 to the actual written seasons
bikes.season = bikes.season.map(season_dict)

# changing the weather column data type to string
```

```
bikes.weather = bikes.weather.astype('str')
# mapping the values to the actual written weathers
bikes.weather = bikes.weather.map(weather_dict)

# writing the final dataframe to an excel file that we will use in our Tableau visualisations. The file will be the 'london_bikes_final.xlsx' file and the sheet name is 'Data'
bikes.to_excel('london_bikes_final.xlsx', sheet_name='Data')
```

10. Screenshots

Figure 1: Moving Average Rides Graph

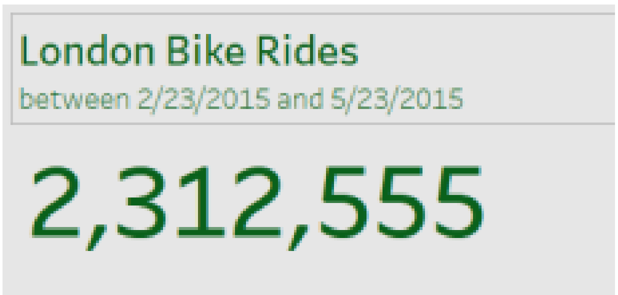

Figure 2: Total Rides Count

Figure 3: Heat Map

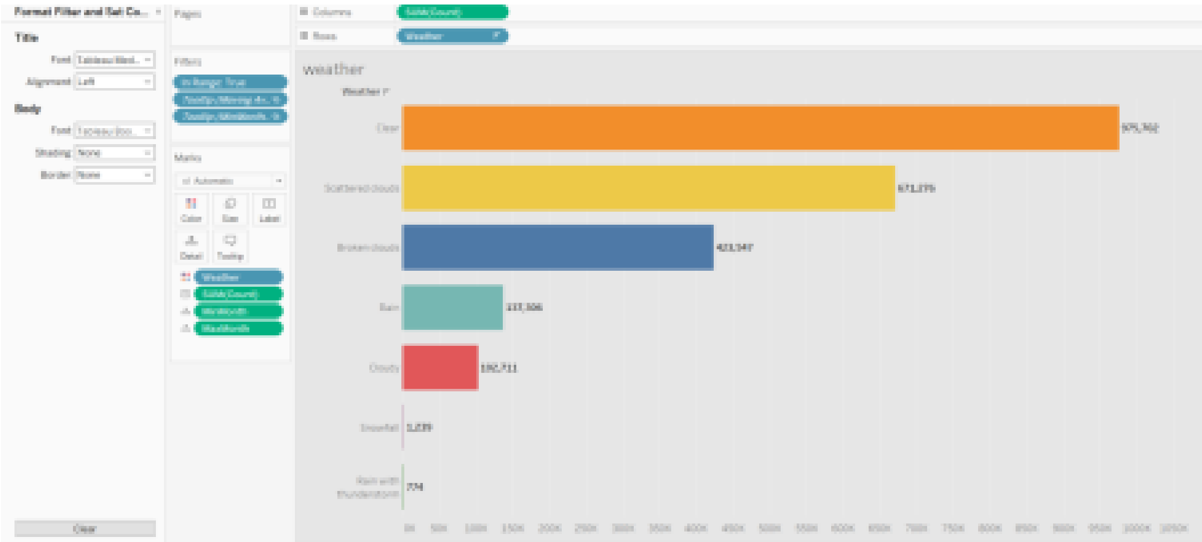

Figure 4: Weather Bar Chart

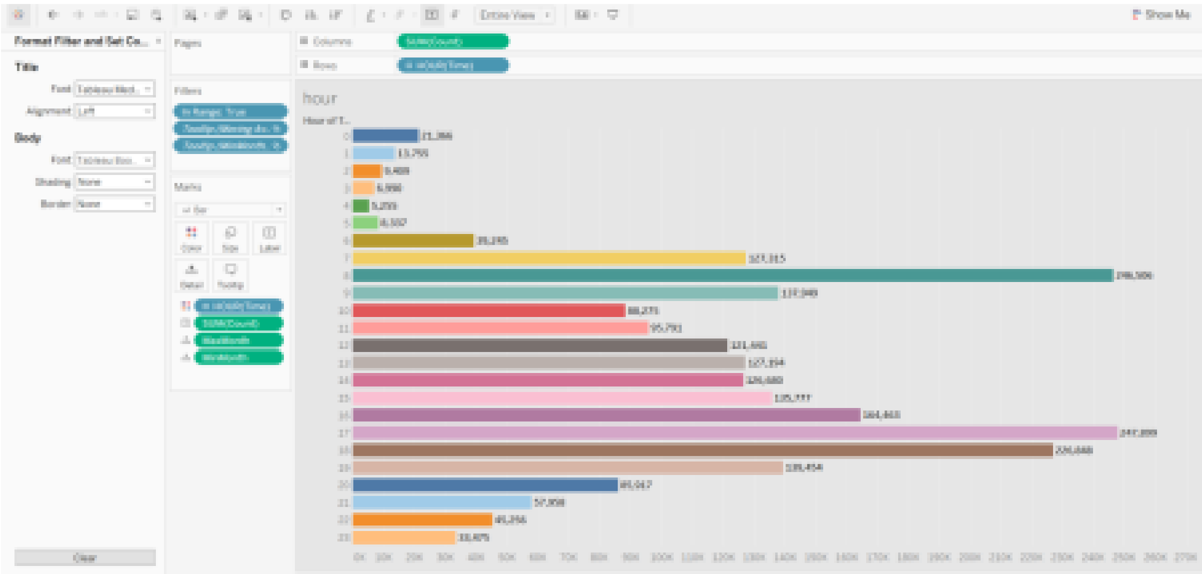

Figure 5: Hour Time Bar Chart

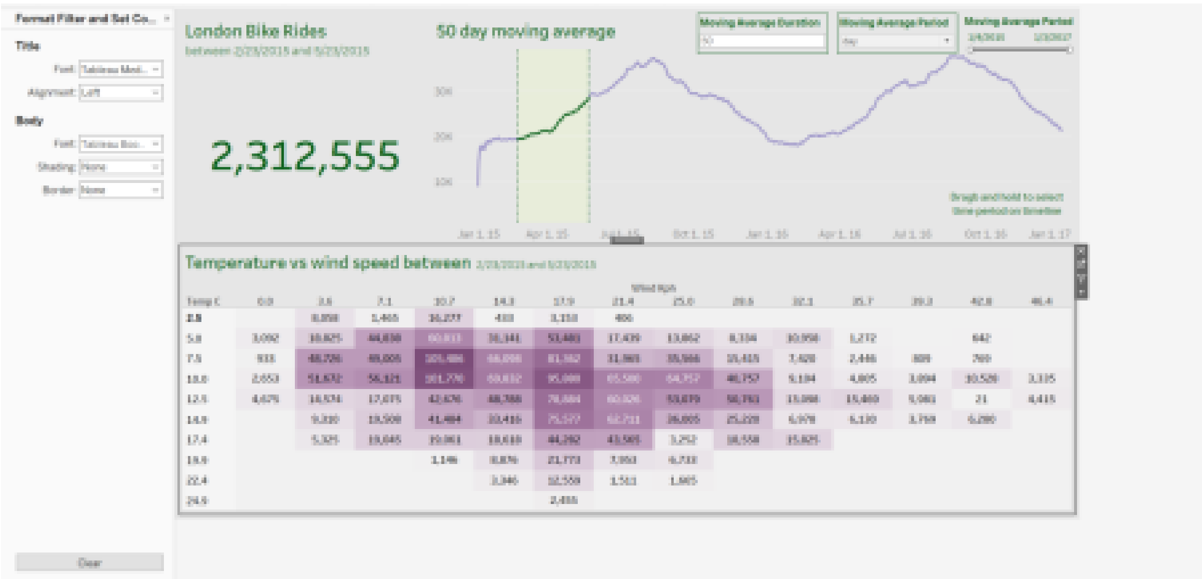

Figure 6: Final Dashboard

11. Conclusion

The dashboard's readability and navigation were the main sources of user experience issues. User testing sessions yielded feedback indicating challenges with locating certain features, comprehending data visualizations, and interacting with the interface. A thorough overhaul of the dashboard design and navigation flow was carried out to solve these issues. An interface that is easy to use and intuitive was created by implementing the principles of user-centered design. To assist users in navigating the dashboard's features, interactive tutorials, contextual tooltips, and clear and simple labeling were incorporated. Usability testing sessions were conducted iteratively, allowing for continual input absorption, and refining of the user experience.