

A Mini Project Synopsis on

Crowd In The Town: Taxi Trip Analysis Using Python

S.E. – Computer Science and Engineering-Data Science

Submitted By

SUMIT SAMANTA	21107003
JANHAVI KASAR	21107030
SANSKRUTI CHAVAN	21107047
SHREYAS REVANKAR	21107065

Under The Guidance Of

PROF.ANAGHA AHER



DEPARTMENT OF CSE-DATA SCIENCE

A.P.SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic Year: 2022-23

CERTIFICATE

This to certify that the Mini Project report on **Crowd In The Town: Taxi Trip Analysis Using Python** has been submitted by Sumit Samanta(21107003), Janhavi Kasar(21107030), Sanskruti Chavan(21107047) and Shreyas Revankar(21107065) who are a Bonafede students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfillment of the requirement for the degree in **Computer Science and Engineering(Data Science)**, during the academic year **2022-2023** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Ms. Anagha Aher
Guide

Prof. Anagha Aher
Head Department of CSE-Data Science

Dr. Uttam D.Kolekar
Principal

External Examiner(s)

- 1.
- 2.

Place: A. P. Shah Institute of Technology, Thane
Date:

TABLE OF CONTENTS

1. Introduction.....	1
1.1. Purpose.....	2
1.2. Objectives.....	3
1.3. Scope.....	4
2. Problem Definition.....	6
3. Proposed System.....	8
3.1. Features and Functionality.....	8
4. Project Outcomes.....	10
5. Software Requirements.....	11
6. Project Design.....	12
7. Project Scheduling.....	17
8. Conclusion.....	20

Acknowledgement

This project would not have come to fruition without the invaluable help of our guide Prof. Anagha Aher. Expressing gratitude towards our HoD, **Prof. Anagha Aher**, and the Department of Computer Science and Engineering-Data Science for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teacher Prof.Ms. Anagha Aher who gave us her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.

Chapter 1

Introduction

Exploratory Data Analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. It helps to understand the underlying patterns and relationships between variables identify outliers for further analysis. New York City is renowned for its iconic taxis, which are an essential mode of transportation for millions of residents and visitors. Every day, millions of people rely on these taxis to get around the city, making them an integral part of New York's transportation system. With the advent of technology and the availability of vast amounts of data, it has become possible to analyze the trips taken by these taxis and gain insights into the patterns and trends of taxi usage in the city.

To better understand the state of the taxi industry in New York City, this report presents an analysis of over 1 billion taxi rides of 2016. We have explored various aspects of taxi usage in the city, including trip duration, passenger counts, and pickup and drop-off locations. By analyzing this data, we hope to gain a better understanding of the factors that influence taxi usage in New York City and provide insights that can inform transportation policy and urban planning. Additionally, we aim to understand the impact of ride-sharing services on the taxi industry and explore ways in which taxis can remain relevant and competitive in the changing transportation landscape.

The process of EDA involves several steps, including:

1. Data collection and cleaning:

Collecting the data from different sources and cleaning it by removing any outliers, missing values, or inconsistencies.

2. Visualization:

Plotting the data using various visualization techniques such as histograms, box plots, scatter plots, etc., to identify patterns, trends, and relationships.

3. Iteration:

Repeating the above steps as needed to refine the analysis and generate new insights.

Overall, EDA is an essential step in any data analysis project as it helps to uncover insights and patterns that may not be apparent through other means. Analyzing taxi trips is an interesting way to understand the transportation patterns in a city. In our project we will explore a dataset of taxi trips in New York City and analyze various aspects of the trips using Python. With the help of various libraries and tools available in Python, we can gain a deeper understanding of the transportation patterns in a city and explore various other aspects of the dataset.

A typical Taxi company faces a common problem of determining the Efficiently assigning the cabs to the passengers during peak demand periods is challenging.

To identify popular pickup and drop off locations in a city. Hence, our project aim is to provide cab company the analysis to anticipate which the popular pick up and drop offs point in the city of New York are so as to increase their rides. They will be able to have a count on total number of passengers travelled and rides complete.

The scope of the project is: -

1. Can be useful for daily commuters to help them reach on time.
2. Can be analyzed by Cab Company to increase their rides at specific time and place.
3. Can be helpful to the cab driver to calculate his total duration of trips.
4. Can be used in the field of trip planning.
5. Can be beneficial for the cab company to keep insights on total counts of passenger and number of trips as per year, month, and week.

1.1 Purpose:

The main purpose of taxi trip analysis is to extract meaningful insights and patterns from data related to taxi trips, such as pick-up and drop-off locations, trip duration and other relevant attribute. By analyzing data on pick-up and drop-off locations, taxi companies can identify areas with high demand and allocate their resources accordingly, leading to better service for customers. Also the company can analyze data on trip

duration, traffic conditions, and other factors, to which taxi companies can optimize their routes and reduce travel time, leading to more efficient operations and cost savings.

Overall, taxi trip analysis can provide valuable insights that help taxi companies, governments, and other stakeholders make data-driven decisions to improve transportation services and enhance the overall customer experience.

1.2 Objectives:

The objectives of taxi trip analysis can vary depending on the specific needs and goals of the organization or individual performing the analysis. However, some common objectives of taxi trip analysis include:

1. For identifying trends and patterns:

Taxi trip analysis can help identify trends and patterns in customer behavior, such as popular pick-up and drop-off locations, peak hours of demand, and customer preferences.

2. To improving operational efficiency:

By analyzing data on trip duration, total passenger count and other factors, taxi companies can optimize their operations to reduce travel time, improve customer service, and reduce costs.

3. To Enhance customer experience:

Taxi trip analysis can help taxi companies understand their customers better and provide more personalized services.

4. Increasing revenue:

Taxi trip analysis can help taxi companies identify new revenue streams, such as offering premium services, partnering with other businesses, and expanding their service areas.

5. Market research:

Taxi trip analysis can help businesses and organizations understand the market demand for taxi services, identify potential customers, and develop targeted marketing strategies.

6. Understanding demand and supply:

By analyzing taxi trip data, one can identify patterns of demand and supply for taxis in different areas and at different times. This information can help taxi companies optimize their fleet and allocate their resources more effectively.

7. Improving flow:

Taxi trip data can also provide insights into the dataset flow which can have a significant impact on the quality of life for residents.

8. Supporting research:

Taxi trip data can be used for research purposes, such as studying the impact of transportation policies or analyzing the behavior of riders and drivers. This can help policymakers make informed decisions and improve the overall quality of transportation services.

Overall, taxi trip analysis can provide valuable insights and help businesses and organizations make data-driven decisions to improve their services, comply with regulations, and achieve their goals, optimizing taxi services to improving traffic flow, enhancing safety and security, and supporting research.

1.3 Scope:

The scope of taxi trip analysis can vary depending on the specific needs and objectives of the organization or individual performing the analysis. However, some common areas of focus within the scope of taxi trip analysis include:

1. Pick-up and drop-off locations:

Analyzing the frequency and distribution of pick-up and drop-off locations can help taxi companies identify areas of high demand and optimize their operations to serve customers more efficiently.

2. Trip duration and route optimization:

Analyzing trip duration and other factors can help taxi companies optimize their routes and reduce travel time, leading to more efficient operations and cost savings.

3. Customer behavior:

Analyzing data on customer preferences, demographics, and other factors can help taxi companies understand their customers better and provide more personalized services.

4. Supply and demand:

Analyzing data on the supply of taxis and the demand for services can help taxi companies adjust their supply to meet demand and avoid oversupply or undersupply.

5. Seasonal trends:

Analyzing data on seasonal trends in taxi demand can help taxi companies prepare for high-demand periods and adjust their operations accordingly.

Overall, the scope of taxi trip analysis can be broad and cover various aspects of taxi services, including customer behavior, service quality, driver behavior, and supply and demand. Analyzing these factors can help taxi companies optimize their operations, improve customer satisfaction, and increase profitability. By leveraging these datasets, taxi companies can make data-driven decisions that improve their operations and provide better services to their customers.

Chapter 2

Problem Definition

Taxi trip analysis is an essential tool for taxi companies and researchers to extract valuable insights from the vast amounts of data generated by taxi services. The problems that can arise while analyzing taxi trip data can be due to a variety of factors. One of the main factors is the complexity and size of the dataset. Taxi trip datasets can be very large, with millions of records, making it challenging to process and analyze the data efficiently. It requires careful consideration and planning to ensure that the analysis is accurate, reliable, and provides valuable insights for taxi companies. However, several problems can arise during taxi trip analysis that can affect the quality and accuracy of the insights obtained.

These problems include:

1. Data quality:

The quality of data used in taxi trip analysis can be a significant issue. Incomplete, inaccurate, or inconsistent data can lead to incorrect conclusions and poor decision-making.

2. Data complexity:

The large volume of data generated by taxi services can make it challenging to analyze and extract meaningful insights. Data preprocessing and analysis techniques may be required to extract valuable insights from the data.

3. Bias:

Bias can be introduced in taxi trip analysis due to the data collection method, or analysis techniques used. Bias can lead to incorrect conclusions and poor decision-making.

Overall, conducting taxi trip analysis requires careful consideration of data quality, privacy and security, complexity, bias, and regulatory compliance to ensure accurate and actionable insights. Taxi companies need to efficiently assign cabs to passengers to provide timely and reliable transportation services while optimizing their operational costs. This can be a challenging task, especially during peak demand periods or in areas with high traffic congestion.

To address this problem, taxi companies can leverage data analysis to better understand patterns of demand and supply for their services. For example, by analyzing historical trip data, they can identify the busiest times and areas of the city and adjust their fleet allocation accordingly.

Chapter 3

Proposed System

A proposed system for taxi trip analysis would involve several components, including data collection, preprocessing, analysis, and visualization. The first step in taxi trip analysis is to collect data on taxi trips. This data may include information on trip start and end times, pickup and drop-off locations. Once the data has been collected, it may need to be cleaned and processed to remove errors, inconsistencies, and missing values. Preprocessing may also involve data transformation, such as converting raw data into a format suitable for analysis.

The next step is to perform various analysis on the preprocessed data to identify patterns, trends, and relationships between variables. The results of the analysis can be visualized using various techniques, such as charts, graphs, and maps, to provide a clear and intuitive representation of the data. This may involve refining data collection methods, improving preprocessing techniques, and updating the analysis and visualization tools.

Overall, a proposed system for taxi trip analysis should be scalable, adaptable, and easy to use, to allow taxi company to extract valuable insights from the data and make data-driven decisions.

3.1 Features and Functionality: -

A taxi trip analysis project typically involves analyzing data from taxi rides to gain insights into various aspects of the rides, such as travel patterns, customer behavior, and driver performance. The features and functionality of such a project can vary depending on the specific objectives and requirements, but some common ones are:

1. Predictive analysis:

It will include predictive analysis to anticipate future trends and patterns based on historical data.

2. Visualization:

The analysis will help to visualize any patterns or trends in the data which involve plotting the data on a map of New York to see which areas have the highest demand for ride services.

3. Efficient services:

Will help ride-hailing services optimize their operations and provide more efficient services to their customers.

4. Route Analysis:

Analyze the most common routes taken by passengers and identify shortest or fastest routes between the different locations.

5. Clustering:

Cluster the pickup and drop off locations which can help identify popular places where there is a high demand for taxis at peak time.

6. Identifying Popular Pickup and Drop-Off Locations:

By analyzing pick-up and drop-off location datasets, taxi companies can identify the most popular locations where passengers are picked up and dropped off. This information can help companies allocate their drivers and vehicles more effectively and improve their overall efficiency.

Overall, the features of taxi trip analysis can provide valuable insights into various aspects of the taxi industry. By leveraging these features, taxi companies can improve their services, reduce costs, and stay competitive in an increasingly crowded market

Chapter 4

Project Outcomes

Project outcomes are the changes that occur as a result of your actions. These typically involve improvements for a product or service. When designing a project, it's important to know what your project outcomes are so you have a way of measuring your success and understand what your overall goal is. They are also valuable for businesses because they help create deliverables to meet their purpose and goal.

There are many possible outcomes that could be gleaned from an analysis of New York City taxi trip data, depending on the specific questions being asked and the methods used to analyze the data. Here are a few potential outcomes:

1. Insights into passenger behavior:

This analysis of taxi trip data can reveal patterns in passenger behavior, such as popular pickup and drop-off locations, number of passenger counts, trip occurred, and common trip distances. This information could be useful to inform decisions about service offerings and marketing strategies.

2. Optimization of taxi routes:

By analyzing data on popular destinations, and other factors, taxi companies can identify ways to optimize their routes and this could lead to more efficient operations.

3. Economic analysis:

An analysis of taxi trip data could provide insights into the economic impact of the taxi industry, including the number of jobs created and the amount of revenue generated.

4. Demand prediction:

By analyzing past taxi trip data, it may be possible to predict future demand for taxi services, based on factors such as time of day, weather, and special events.

5. Route optimization for accessibility:

By analyzing data on pickup and drop-off locations, taxi companies could identify areas that are poorly served by public transportation and adjust their routes to better serve these areas.

Chapter 5

Software Requirements

Software Requirement Specification (SRS) Format as name suggests, is complete specification and description of requirements of software that needs to be fulfilled for successful development of software system. These requirements can be functional as well as non-functional depending upon type of requirement. This is description of features and functionalities of the target system. Requirements convey the expectations of users from the software product. It is the responsibility of system analyst to document the requirements in technical language so that they can be comprehended and useful by the software development team.

The software and technology stacked used in development of New York taxi trip analysis are:-

- Power Bi Desktop 2.116.884.0
- Pycharm Community Edition 2022.1.3
- Mysql 8.0 CE

Chapter 6

Project Design

Design is the first step in the development phase for any engineering product (or) system. It may be defined as “the process of applying various techniques and principles for the purpose of defining a device, a process, or a system insufficient detail to permit its physical realization”. Software design is an iterative process through which requirements are translated into a ‘Blue print’ for constructing the software.

The design is represented at a high level of abstraction, a level that can be directly translated to specific data, functional and behavioral requirements. The interface design describes how the software communicates within itself, to systems that interoperate with it, and with humans who use it.

1) Login Page

This is the very first page of the application. Here the user needs to enter his/her username and password to sign up in the application. If he/she is new then they need to first register and the create a new account.

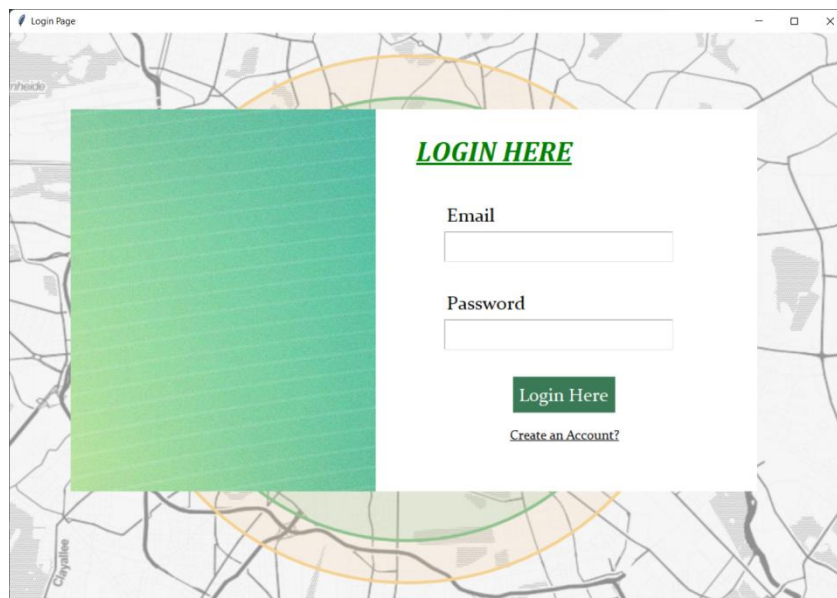
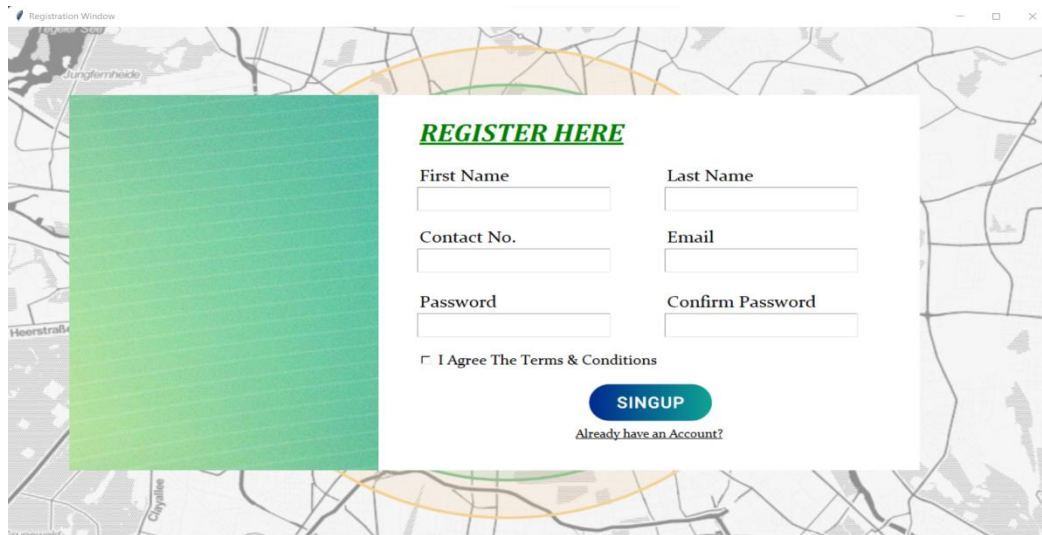


Figure 6.1 Login Page

2) Registration Page

If the user is a new user he/she needs to first create an account by entering the personal details such as name, email, password, contact no etc. and then login.



The screenshot shows a web browser window titled "Registration Window". The background is a map of a city. Overlaid on the map is a white registration form. The form has a green header with the text "REGISTER HERE". Below the header, there are six input fields arranged in two columns: "First Name", "Last Name", "Contact No.", "Email", "Password", and "Confirm Password". Below the input fields, there is a checkbox labeled "I Agree The Terms & Conditions". At the bottom of the form, there is a blue button labeled "SIGNUP" and a link labeled "Already have an Account?".

Figure 6.2 Registration Page

3) Main Dash Board

This is the first dashboard of the project. Here the user will be able to find the main analysis and answers to the questions he/she is in search of. It contains the major statistics such as the total passenger count, total trips occurred, pickup and drop-off districts as well as the regular and holiday passenger counts.

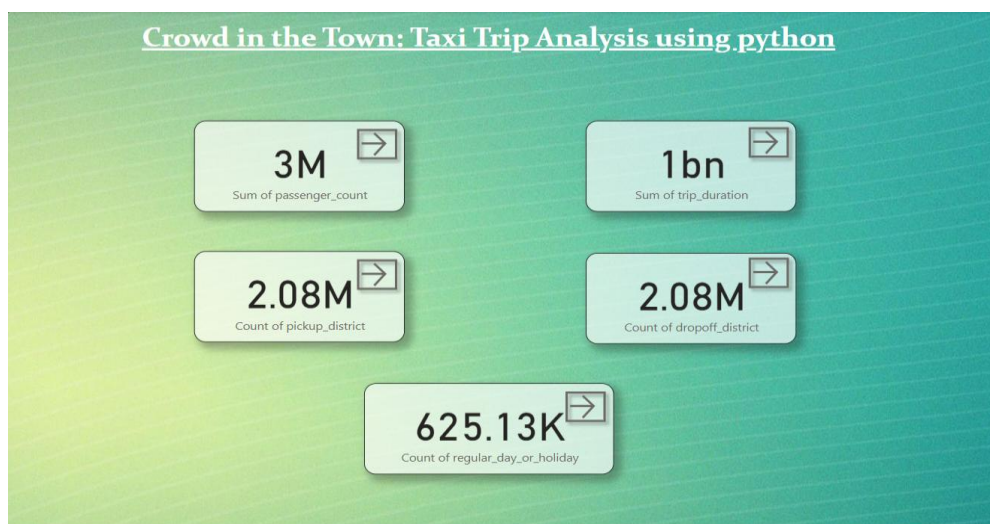


Figure 6.3 Main Dashboard

4) Analysis of total passenger count

This is the next page where the user will be able to find all the analysis regarding the passenger count. This page will provide the passenger count by period of a day, weeks, months, pickup and drop-off district.

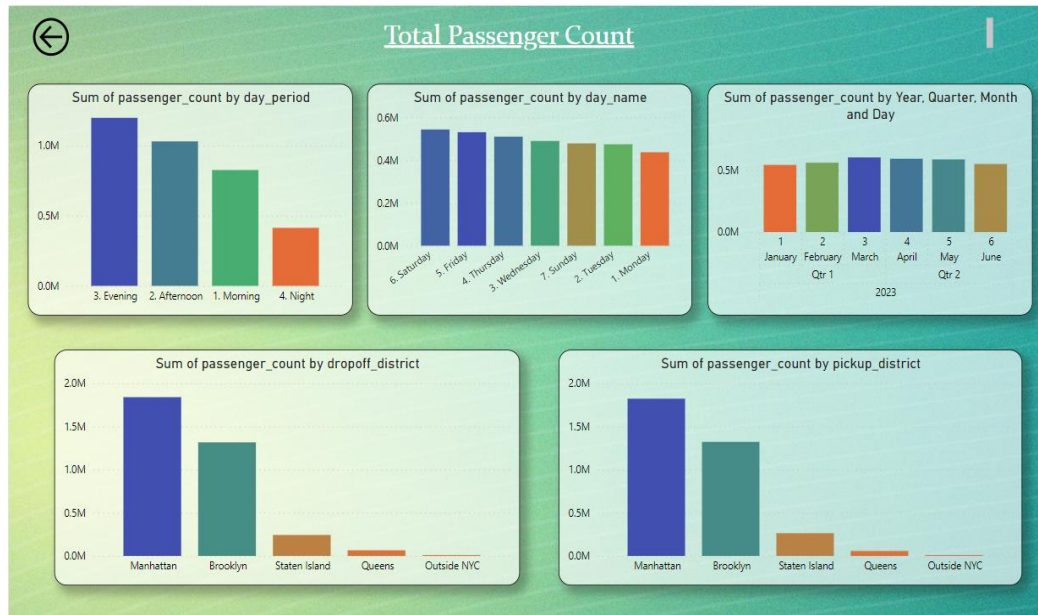


Figure 6.4 Passenger count analysis

5) Analysis of total trip count

This is the next page where the user will be able to find all the analysis regarding the total trip counts. This page will provide the trip count by period of a day, weeks, month seasonal trends as well as weekend and weekdays analysis.



Figure 6.5 Total Trip count analysis

6) Analysis on pickup locations

This is the next page where the user will be able to find all the analysis regarding the pickup locations. This page will provide the New York map based on pickup latitude and longitude, trips occurred at the places based on pick up hour and neighborhood districts.

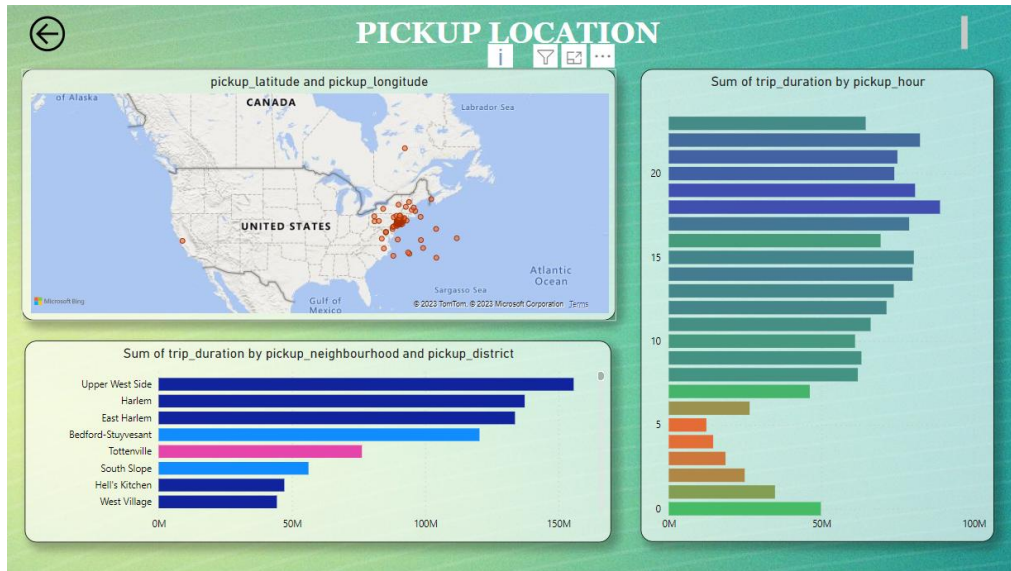


Figure 6.6 Pickup locations analysis

7) Analysis on drop-off locations

This is the next page where the user will be able to find all the analysis regarding the drop-off locations. This page will provide the New York map based on drop-off latitude and longitude, trips occurred at the places based on drop off hour and neighborhood districts.

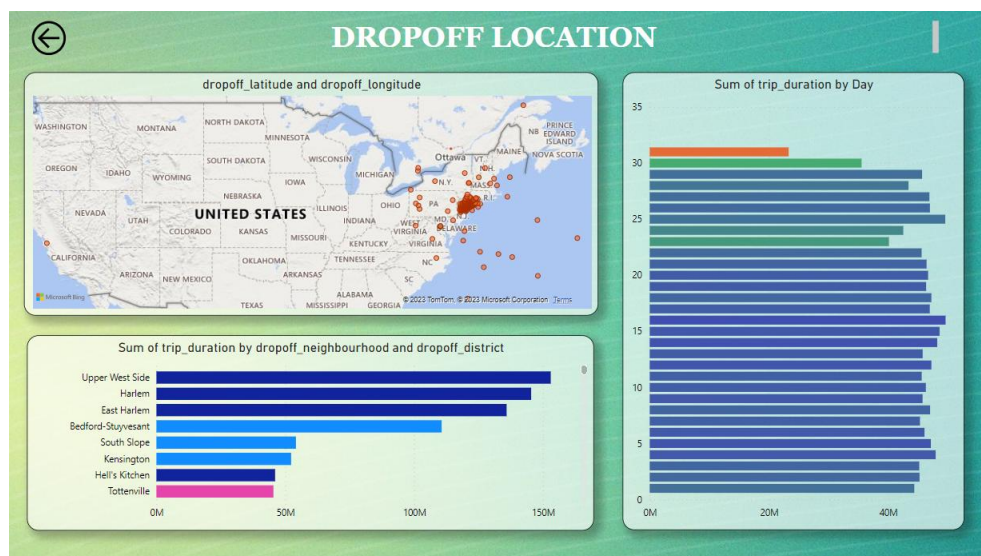


Figure 6.7 Drop-off locations analysis

8) Analysis of trip count on basis of weekend/weekdays/holidays

This is the next page where the user will be able to find all the analysis regarding the total trip counts. This page will provide the trip count by weekend/weekday and regular days/holidays.

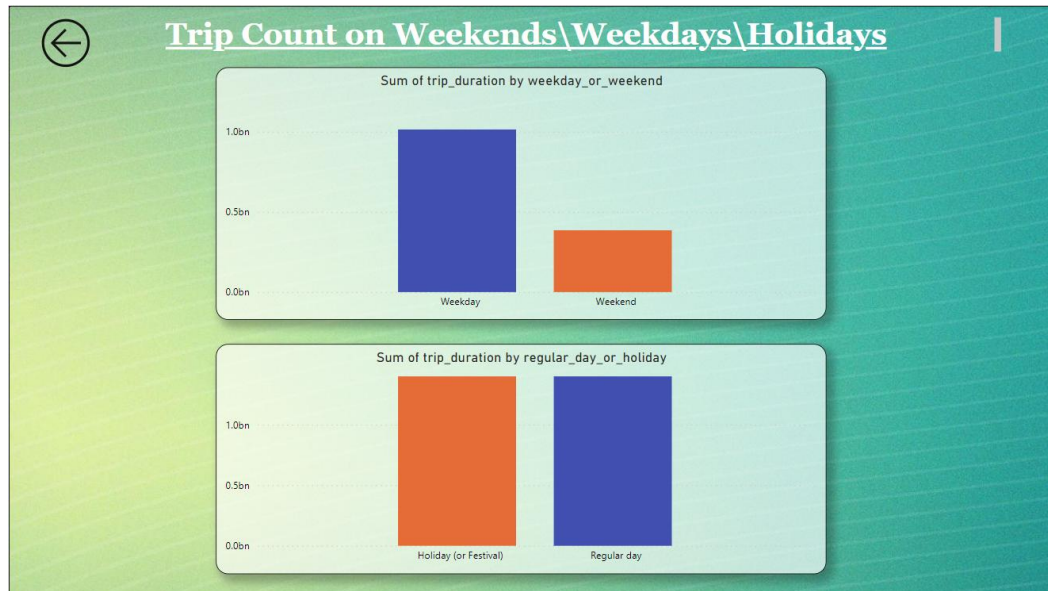


Figure 6.8 Total Trip count analysis by weekend/weekdays/holidays

Chapter 7

Project Scheduling

In project management, a schedule is a listing of a project's milestones, activities, and deliverables. Usually, dependencies and resources are defined for each task, then start and finish dates are estimated from the resource allocation, budget, task duration, and scheduled events. A schedule is commonly used in the project planning and project portfolio management parts of project management. The development and maintenance of the project schedule is the responsibility of a full-time scheduler or team of schedulers, depending on the size and the scope of the project. The project schedule is a calendar that links the tasks to be done with the resources that will do them. It is the core of the project plan used to show the organization how the work will be done, commit people to the project, determine resource needs, and used as a kind of checklist to make sure that every task necessary is performed.

A Gantt chart is a type of bar chart that illustrates a project schedule. Modern Gantt charts also show the dependency relationships between activities and the current schedule status. This chart lists the tasks to be performed on the vertical axis, and time intervals on the horizontal axis. The width of the horizontal bars in the graph shows the duration of each activity. Gantt charts illustrate the start and finish dates of the terminal elements and summary elements of a project. Terminal elements and summary elements constitute the work breakdown structure of the project. Modern Gantt charts also show the dependency (i.e., precedence network) relationships between activities. Gantt charts can be used to show current schedule status using percent-complete shadings.

Here in the figures 7.1 the rows of the chart contain the task titles such as the project conception and initialization as well as the project design and implementation which in subdivision contains the group formation, topic finalizing, prototype, GUI designing, backend implementation etc. The columns contain the duration of the task completed, percentage of work completed, number of weeks required to complete a particular task, the specific dates, the team members who contributed towards the completion of task. The detailed explanation of the Gantt chart is as explained below: -

The major task of project conception and initiation was executed by the July month end around 10/01/2023. The task of initiation included many more sub tasks such as the group formation and topic finalization which was performed during the 1-2 weeks of project initialization (10/01/2023-24/01/2023). The group formed included 4 members Sumit Samanta, Sanskruti Chavan, Janhavi Kasar, Shreyas Revankar and the finalized topic was Crowd in the Town- Taxi Trip Analysis Using Python. Further the upcoming weeks (10/01/2023-24/01/2023) led to the task of identifying the scope and objectives of the mini project. The next sub task was to identify the functionalities of the project which was done by the two members Sumit Samanta and Janhavi Kasar in a span of one week from 10/01/2023 to 17/01/2023. The discussing of the project topic with the help of paper prototype was carried out by Sanskruti Chavan, Janhavi Kasar, and Shreyas Revankar in the next one week from 17/01/2023 to 24/01/2023. The next main task of Graphical User Interface (GUI) designing was completed by Shreyas Revankar and Sanskruti Chavan within 3 weeks from 24/01/2023 to 07/02/2023. The next 1 week from 14/02/2023 to 06/03/2023 the all members worked on the preparation of presentation review-I. The next major task was of project designing and implementation. It took in all 5 weeks to complete the final implementation. The database Design and connectivity of all modules was done by Shreyas Revankar, Sumit Samanta and Janhavi Kasar during the course time of 1 weeks from 06/03/2023 to 14/03/2023. The integration of all modules and report writing was completed by Sanskruti Chavan and Janhavi Kasar during 21/03/2023 to 28/03/2023. The preparation of final presentation II work was equally shared by all the group members in the time of 2 weeks from 04/04/2023 to 20/04/2023.

Chapter 8

Conclusion

In conclusion, a taxi trip analysis project using a dataset in Python can provide valuable insights into taxi operations and customer behavior. By analyzing key features such as pick-up and drop-off locations, passenger count, time and date, trip distance, fare amount, and payment method, taxi companies can make data-driven decisions to improve their services and provide better experiences for their customers. Python is a powerful programming language for data analysis and has many libraries such as Pandas, NumPy, Seaborn and Matplotlib that make it easy to work with large datasets. By using these libraries, data scientists can clean, preprocess, and analyze taxi trip datasets to extract meaningful insights. Some of the common analysis techniques that can be used in a taxi trip analysis project include data visualization, machine learning algorithms, and predictive modeling. By applying these techniques, data scientists can identify patterns, trends, and anomalies in taxi trip data, which can help taxi companies optimize their operations and improve their services. Overall, a taxi trip analysis project using a dataset in Python can provide a comprehensive understanding of taxi operations and customer behavior, which can help taxi companies make data-driven decisions to improve their services, increase efficiency, and enhance customer satisfaction.