



SDAIA T5 bootcamp

Exploratory Data Analysis (EDA)

Course Project Introduction

The Metropolitan Transportation Authority (MTA)

proposal

Basma Ghazi Abdullah Alduaiji

25/09/2021

Contents

| | |
|----------------------|---|
| Table of tables..... | 2 |
| Introduction..... | 3 |
| Backstory | 3 |
| Used data set | 3 |
| Used tools | 4 |
| Conclusion | 4 |
| References..... | 5 |

Table of tables

| | |
|----------------------------|---|
| Table 1 MTA Data Set | 3 |
|----------------------------|---|

Introduction

Due to the extreme growth of population in the USA, the need to transport projects significant to enable people of reaching different destinations in an easier and faster way. As a result, the government of New York City established the Metropolitan Transportation Authority (MTA) transport project in 1904 [1], and it has used until now. However, the MTA project still suffering from multiple issues such as the crowded train stations and the long waiting time and lack of arrival time accuracy in some cases. Hence, the MTA needs some improvement to increase passengers' satisfaction levels. So, this project aims to provide a solution to overcome MTA issues. At the same time, data science can be utilized to find the best solution.

Backstory

As a well-known consultant company working in data science, which provides consulting and solutions throughout analysing data. The last week, the company has received an email from a government institution. The institution works on improving New York City. The email contains a proposal to provide a solution to organize the overwhelmed stations specially during rush hours and solve the other mentioned issues, to reach the required satisfaction levels using data science.

Used data set

This project will use the MTA data set which is an open source and freely accessible data set. Moreover, the MTA data set contains multiple fields with different data types as show in table 1.

Table 1 MTA Data Set

| | |
|----------|---|
| C/A | Control area of each turnstile. |
| UNIT | Remote unit for the station. |
| SCP | Subunit Channel Position represents a specific address for a turnstile. |
| STATION | Represents the station name. |
| LINENAME | Represents the line name in the station. |
| DIVISION | Represents the line that the station belonged to it. |
| DATE | Represents the date when data acquiring. |
| TIME | Represents the time when data acquiring. |

| | |
|---------|--|
| DESC | Represents the type of data either regular or recover. |
| ENTRIES | Represents the cumulative enters. |
| EXITS | Represents the cumulative exit. |

However, The Individual sample from MTA data set are three months: May, June, and July in year of 2021. Moreover, the expected characteristics used to find a solution are the attributes from MTA data set by company such as Time, Date, Station and Entries. Furthermore, the predicted target is to get the most crowded stations in a specific date or time From the MTA data set as well as the number of divisions.

Used tools

The company will utilize the **tools** provided as a module in python such as Pandas, Matplotlib, Seaborn NumPy, Word Cloud and much more. The python libraries support multiple data analysis and cleaning methods to ensure the data are clean and ready to visualize. Furthermore, SQLalchemy can provide the ability to write the SQL queries and get specific data according to query. For that reason, python module tools are sufficient, and the company won't need or use any other additional tools.

Conclusion

The provided solution from the company is to increase the number of stations near to the most crowded ones. Moreover, the company aims to implement the solution by exploring the data set then cleaning it. After that, the company will use the Python Seaborn and Matplotlib module to visualize data within different chart such as pie chart. Finally, the company will write some queries which represent data based on some cases.

References

[1] En.wikipedia.org. 2021. *New York City Subway - Wikipedia*. [online] Available at: <https://en.wikipedia.org/wiki/New_York_City_Subway> [Accessed 25 September 2021].