TUBE TWIN (PASSENGER COUNT FORECASTING /GENERAL TUBE ANALYSIS)REPORT

EXECUTED BY INTERN TEAM-6 2201

TEAM MEMBERS

Fielami Emmanuel David
Michael Mamah
Hakim Balogun
Kelvin Mwaniki
Harmony Odumuko
Endurance Arienkhe

PROJECT PROPOSERS

Tariq Desai
Paul Dos Santos

SUPERVISORS

Elizabeth Stanley
Zintle Faltein

1. PROJECT SUMMARY

With around 400km of rails, over 267 stations, and more than 1.3 billion passenger journeys each year (according to the project data), quickly and safely moving passengers through stations and onto trains is an ongoing priority for the London Underground. Gaining deeper insights into passenger counts and traffic at the different stations will help Transport for London (TfL) make better decisions and plan for future operations.(Li-Yang (Edward) Chiang et al., 2017).

This project's objective is to develop a digital twin of the tube network system. This system will provide a graphical network, a forecasting model for passenger count, and tube analyses for recommendations on operations and traffic flow. Analyses will be based on historical data collected from TfL's open data API, which shows records of the number of people entering and exiting every station in fifteen-minute increments.

Our team was able to build a web-based data application with streamlit where we integrated the systems we created into different pages. These pages include one for the graphical representation of the tube network, another for the forecasting model to predict the count of passengers across different stations of the tube, and one that serves as a dashboard for the exploratory data analysis (EDA). Also, we have designed a simulation of the network between the tube's top stations.

2. INTRODUCTION & PROBLEM STATEMENT

Transport for London (TfL) is a local government body responsible for most of London's transport network, including the London Underground (LU, aka The Tube). The tube is a public railway transit system that serves Greater London and its nearby counties.

London has a population of approximately 9 million people, on top of which hundreds and thousands of tourists visit it every day. This presents some challenges to TfL concerning operations of the tube railway network which includes the following:

- Network operations and surveillance.
- Traffic flow, analyses, and control.
- Infrastructure management and maintenance.

Although not limited to these, the above are the main focus which, all together, form the problem we tried to solve.

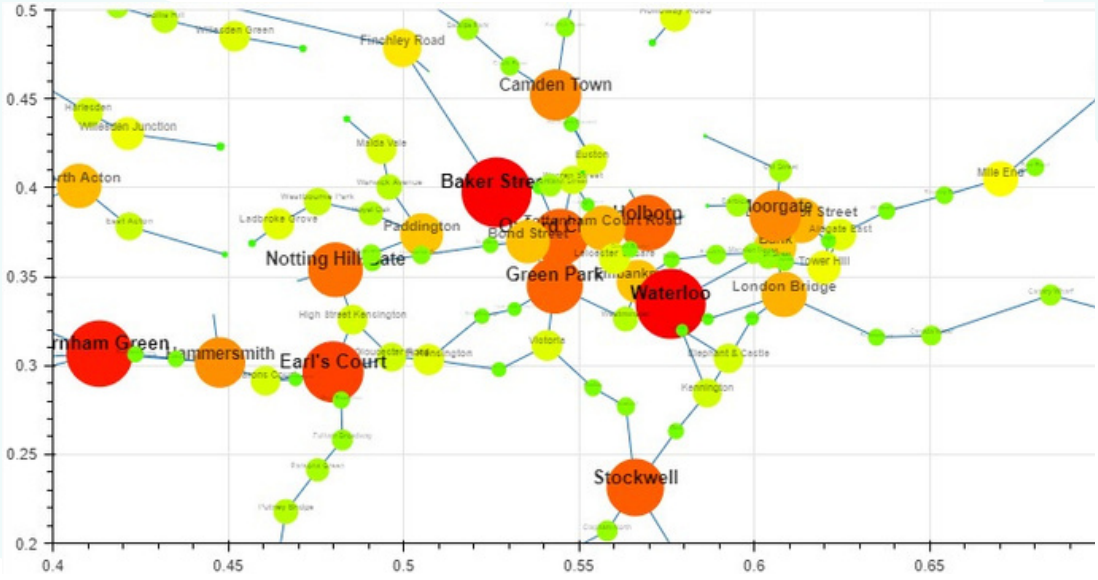
3. OUR SOLUTION

As a solution to the problems mentioned above, our team created a few systems designed to address each of them. These include the following:

3.1 **A Graphical Network of the Tube**

We generated a graph representation of the tube network using NetworkX. From this graph, we were able to make a few analyses like determining the important and busiest stations within the network. Importance is weighted based on nodes connection, while usage is based on edges connections.

Below is a generated graph showing stations based on their importance. Red nodes are more important and green ones are the less important ones.



3.2 **A Simulation of the Tube Network**

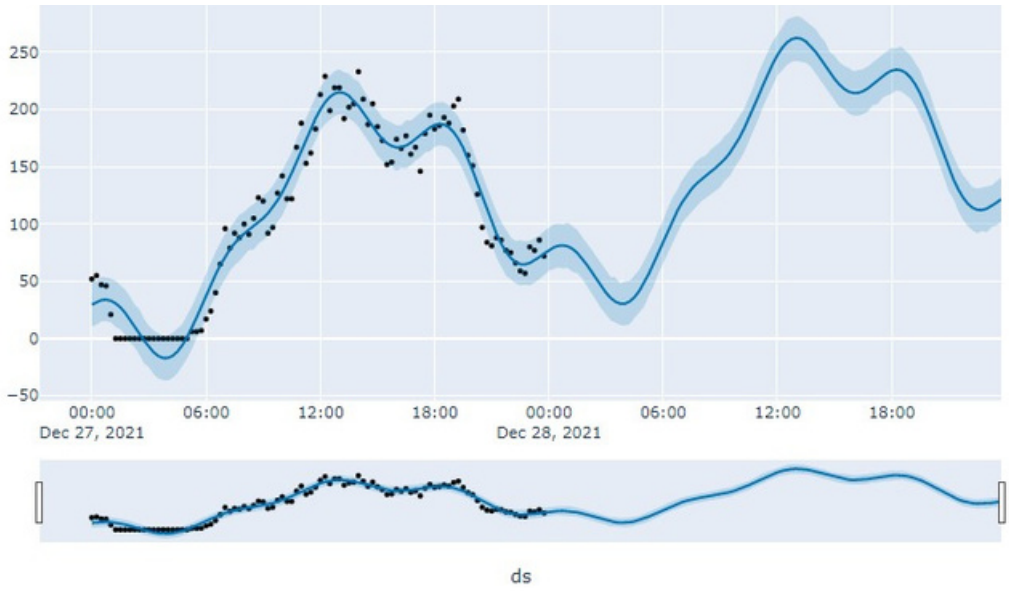
We created a simulation of the tube network using SUMO. For this, we selected a few stations based on their importance and usage ranking.

3.3 **A Model to Forecast Passenger Count**

We created a time-series and several multi-target regression-based models to forecast passenger count based on stations, year, day and other schedule metrics. We achieved this by making use of the Python Sklearn library and FBProphet. This will allow Tfl to predict traffic and plan ahead of time.

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

Prophet is open source software released by Facebook's Core Data Science team. It is available for download on CRAN and PyPI.(Taylor SJ and Letham B, 2017).



Forecast image of the Moorgate station OUT direction on Saturdays

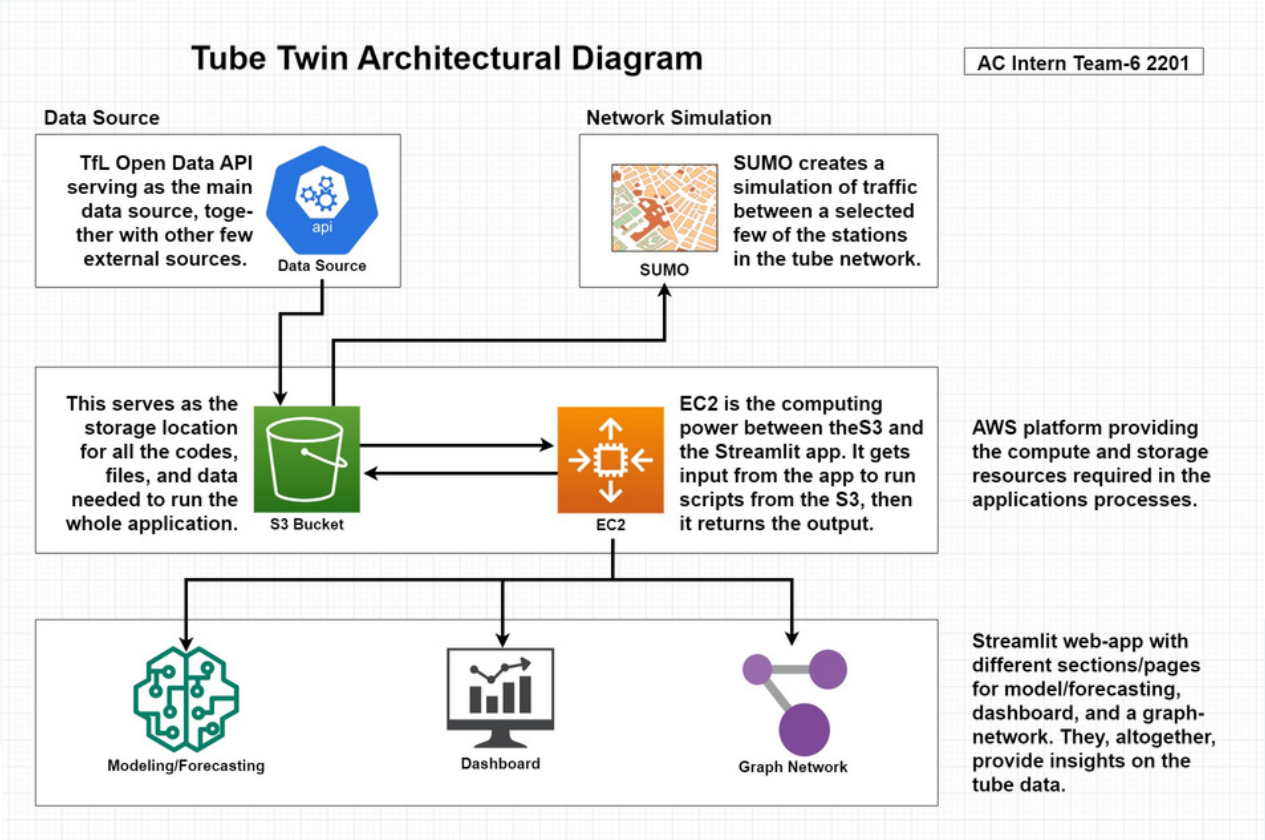
3.4 **A Dashboard for Traffic Analyses**

We were able to create a dashboard where we can generate analyses and gain insights from the results. This will allow for better decision-making and also plan properly for the tube operations.

4. METHODS AND PROCESSES

All applications were developed using python programming language and several open software packages like NetworkX, FBProphet, SUMO, and Streamlit.

The diagram below depicts the flow of processes of the project applications from start to finish. These include the use of AWS resources (EC2 and S3) for computing and storage resources.



The process starts when data is collected from the sources, cleaned and shaped into the required formats. They are then saved into the bucket on AWS where our applications will perform specified tasks on them. All application codes are also stored in the bucket where the EC2 will run them when it gets triggered by our Streamlit app.

5. CONCLUSION

The Tube Twin analysis and and passenger forecasting conducted in this project provides the solution to the traffic issues in the heavily-weighted Tube station lines. The deployed web application adequately covered the top 14 stations for passenger forecasting, while the general time series data analysis and and the graphical network representation provided an overall analysis. The project outcome gives a reliable prototype system that, if deployed and interactively developed, can support TFL in achieving many of it's Tube management objectives which includes traffic control.

6. RECOMMENDATION

The Tube Twin project by Explore AI (2022) team 6 provides a good ground for further research and improvement of passenger flow and traffic analyses on the London Tube. We recommend expansion in the scope of passenger forecasting (i.e. more stations) taking into account certain social activities and environmental factors (e.g. social events and weather).

The next research on the project should consider a live streaming data collection of the tube for a minimum period of one year as this will allow for collection of comprehensive real-time data for higher forecasting accuracy.

This project and report are limited in scope based on the limited time to deliver a more robust system, as such, this report does not provide a conclusive and thorough explanation of passenger flow on the Tube, neither did it provide all intended solutions. This is because available data of the Tube accessible by the team for this research is not sufficient enough for more robust Tube forecasting solution or problem that may arise.

Lastly, to be able to achieve interactive passenger forecasting system that not only serve the management of the Tube but also the passengers, The Transport for London (TFL) can consider integrating Live Weather Conditions at the point of data generation to enable the analysis and forecasting of passenger counts given a certain weather condition

6. REFERENCES

Listed below are the resource links, references, and data sources used for this project.

- Li-Yang (Edward) Chiang, Robert Crockett, Ian Johnson, Aidan O'Keefe. (2017). Passenger Flow in the Tube.
- Taylor SJ, Letham B. 2017. Forecasting at scale. PeerJ Preprints 5:e3190v2 https://doi.org/10.7287/peerj.preprints.3190v2
- Transport for London. (n.d.). Open data policy. Retrieved April 09, 2017, from https://tfl.gov.uk/info-for/open-data-users/open-data-policy