Imperial College London

COMMUTER TRAVEL BEHAVIOUR ON LONDON UNDERGROUND USING AUTOMATIC CLUSTERING OF OYSTER CARD DATA

T8

Kawthar Aljufairi

Department of Civil and Environmental Engineering
Imperial College London

1. INTRODUCTION

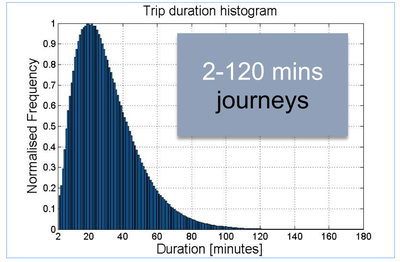
The increase in usage of London Underground is more rapid than the increase in the capacity, which is restricted by physical constraints due to the original design of the tunnels and the length of platforms. This led to TfL seeking other methods to increase the capacity of the system other than physically adding more space. For TfL to decide on a strategy, it needs to have enough evidence that is based on accurate statistical tests of the current usage of London Underground. The introduction of automatic fare collection through Oyster card system makes it possible to track the behaviour of commuters.

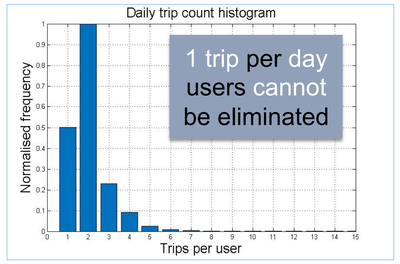
2. DATA

The data available for this research project was collected through Oyster card validation across the network and is provided by the customer experience analytics department of TfL. The data comprises of a two weeks sample from autumn 2013. Travel behaviour of commuters over ten weekdays was analysed.

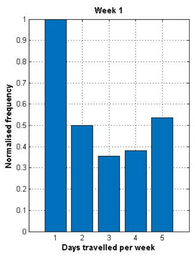
3. DATA REDUCTION

- Commuters are usually defined as those who make the same return trip for 5 days per week.
- London Underground data was statistically tested to reduce the data according to findings.

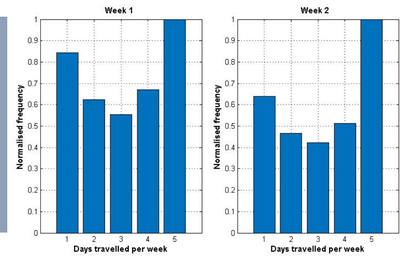




Users who travelled on either weeks



Users who travelled on both weeks

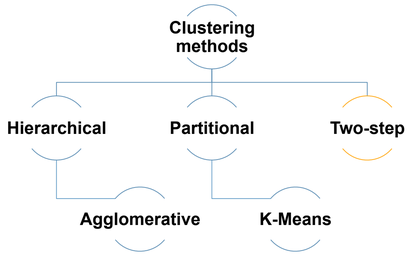


168m data entries

24m data entries (12m commuters)

Commuters are those who consistently travelled for at least 3 days per week for two weeks in a row.

4. AUTOMATIC CLUSTERING



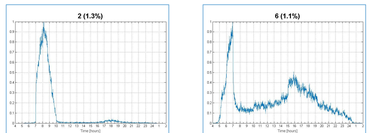
- Two-step clustering was chosen for clustering:
  - It handles very large data sets.
  - Number of clusters is not an input.

Step 1: Pre-cluster the cases (or records) into many small sub-clusters.
Step 2 Cluster the sub-clusters resulting from pre-cluster step into the suitable number of clusters.

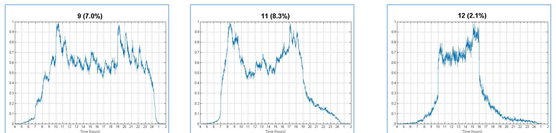
5. RESULTS

14 clusters **68.2%** of commuters

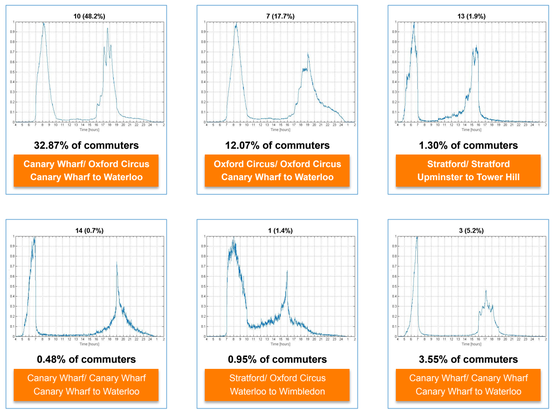
Morning spike behaviour



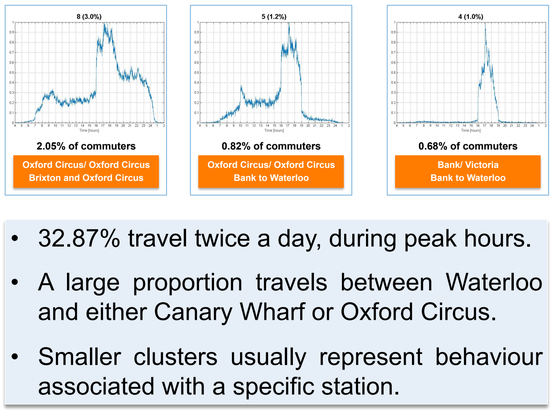
Throughout the day behaviour



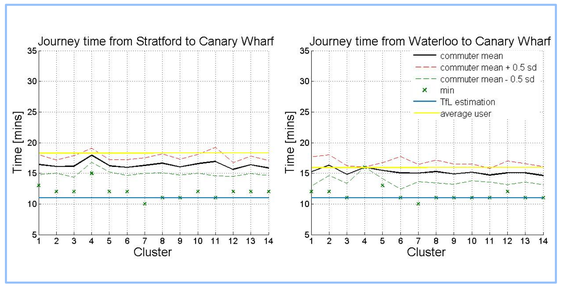
Distinctive two-spike behaviour



Evening spike behaviour



- 32.87% travel twice a day, during peak hours.
- A large proportion travels between Waterloo and either Canary Wharf or Oxford Circus.
- Smaller clusters usually represent behaviour associated with a specific station.



- There is no clear correlation between cluster number and mean travel time.
- Commuters usually have a better journey time experience than the average user.
- TfL underestimates time taken to complete a journey.

6. RECOMMENDATIONS

- TfL could use automatic clustering to identify the type of commuters using LU without collecting personal information from them.
- TfL could utilise this to provide Travelcard plans that better suit the behaviour of commuters taking into consideration both their temporal and spatial behaviour.
- TfL could reduce the volume of people taking a specific route that cannot be improved in capacity by adding extra charges to users taking it at specific times, utilising route validators.

7. ACKNOWLEDGMENTS

- Special thanks to Dr. Majumdar for his supervision and guidance throughout the project.