Uber Movement Data

Problem Statement

Uber travel time data has been used for the below analysis. Objective of the analysis is to Predict approximate time to travel between point A to point B at given interval of the day. We have considered 2019 Quarter4 data for the analysis. Data provided by Uber is aggregated at different levels i.e. monthly, weekly, daily and hourly.

Data Sources and descriptions

- Uber Movement Data:
  - https://movement.uber.com/explore/bangalore/travel-times/query?si=573&ti=&ag=wards&dt[tpb]=ALL_DAY&dt[wd;]=1,2,3,4,5,6,7&dt[dr][sd]=2020-03-01&dt[dr][ed]=2020-03-31&cd=&sa;=&sdn=&lat.=51.51262&lng.=-0.1658002&z.=12&lang=en-US

- Time Frame:
  - 2019 – Quarter 4

- Wards data:
  - Bangalore wards file contains the list of wards with their spatial co-ordinates
  - This is in json format

- Aggregated data:
  - Uber provides the data of average time taken to travel between from point A to point B
  - This data has been aggregated at weekly, daily and hourly data
  - This file is in the excel format

➢ Data Dictionary:

Wards data

Bangalore Wards	
Field Name	**Sample Value**
type	FeatureCollection
features__type	Feature
features__properties__WARD_NO	2
features__properties__WARD_NAME	Chowdeswari Ward
features__properties__MOVEMENT_ID	1
features__properties__DISPLAY_NAME	Unnamed Road, Bengaluru
features__geometry__type	MultiPolygon
features__geometry__coordinates__001	77.5922939
features__geometry__coordinates__002	13.0971962

Aggregated data

Bangalore-wards-2019-4-All-HourlyAggregate	
Field Name	**Sample Value**
sourceid	143
dstid	141
hod	7
mean_travel_time	314.3
standard_deviation_travel_time	197.14
geometric_mean_travel_time	277.05
geometric_standard_deviation_travel_time	1.6

hod -> hour of the day

Bangalore-wards-2019-4-All-WeeklyAggregate	
Field Name	**Sample Value**
sourceid	145
dstid	121
dow	7
mean_travel_time	1334.07
standard_deviation_travel_time	561.3
geometric_mean_travel_time	1268.67
geometric_standard_deviation_travel_time	1.35

dow -> day of the week

Bangalore-wards-2019-4-All-MonthlyAggregate	
Field Name	**Sample Value**
sourceid	27
dstid	196
month	10
mean_travel_time	2705.92
standard_deviation_travel_time	912.98
geometric_mean_travel_time	2565.14
geometric_standard_deviation_travel_time	1.4

Data Preparation steps for wards data:

Step 1: Converted json file to csv

Step 2: Prepared data such there are no blank cells in between the range of co-ordinates (refer below image)

Step 3: Final output file is validated against the original input file

Before cleansing:

type	features_type	features__properties__WARD_NO	features__properties__WARD_NAME	features__properties__MOVEMENT_ID	features__properties__DISPLAY_NAME	features__geometry__type	features__geometry__coordinates__001	features__geometry__coordinates__002
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.5922939	13.0971962
							77.5909447	13.0984219
							77.58908	13.1000567
							77.5873309	13.1013747
							77.5871202	13.1015574
							77.5869043	13.1017267
							77.5865139	13.1020327
							77.5859112	13.1025052
							77.5850652	13.1031732
							77.5841045	13.1040241
							77.5831971	13.1049004
							77.5825264	13.1056134

After cleansing:

type	type	WARD_NO	WARD_NAME	MOVEMENT_ID	DISPLAY_NAME	feature_type	coordinates__001	coordinates__002
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.5922939	13.0971962
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.5909447	13.0984219
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.58908	13.1000567
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.5873309	13.1013747
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.5871202	13.1015574
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.5869043	13.1017267
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.5865139	13.1020327
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.5859112	13.1025052
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.5850652	13.1031732
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.5841045	13.1040241
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.5831971	13.1049004
FeatureCollection	Feature	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	MultiPolygon	77.5825264	13.1056134

Exploratory Data Analysis

Wards data:

Field Name	Sample Value	Unique values
type	object	1
type.1	object	1
WARD_NO	int64	198
WARD_NAME	object	198
MOVEMENT_ID	int64	198
DISPLAY_NAME	object	198
feature_type	object	1
coordinates__001	float64	21300
coordinates__002	float64	21300

All fields are 100% populated
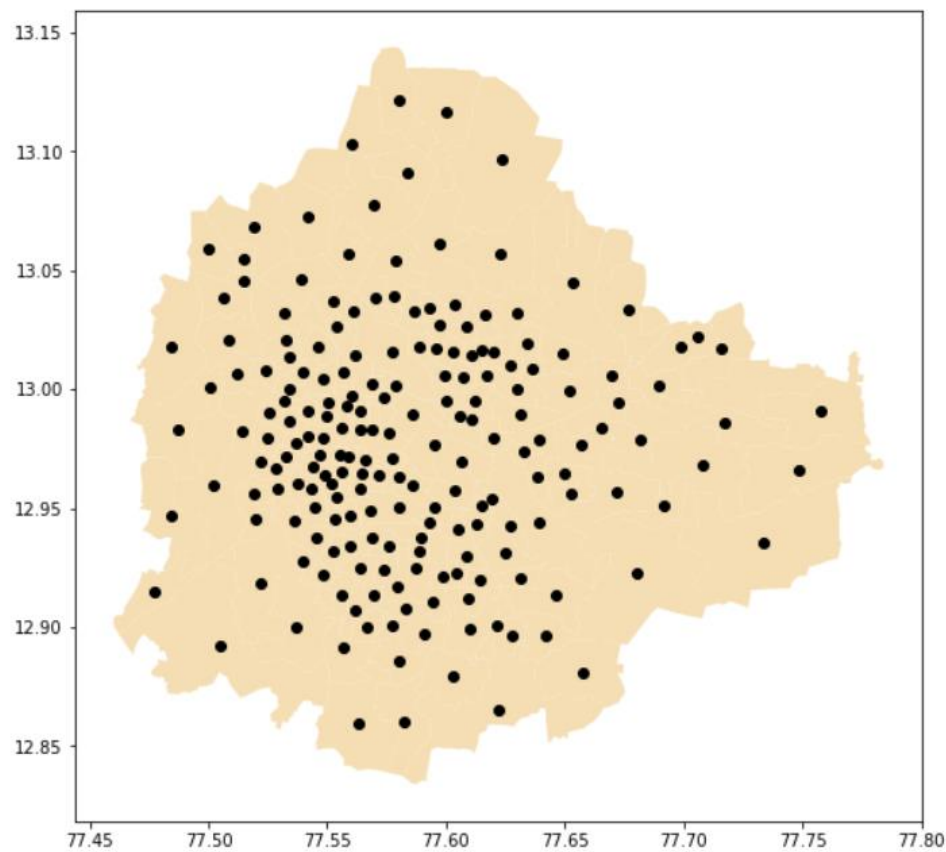
Sample Data:

	WARD_NO	WARD_NAME	MOVEMENT_ID	DISPLAY_NAME	geometry
0	2	Chowdeswari Ward	1	Unnamed Road, Bengaluru	POINT (77.58042 13.12171)
1	3	Atturu	2	9th Cross Bhel Layout, Adityanagar, Vidyaranya...	POINT (77.56004 13.10280)
2	4	Yelahanka Satellite Town	3	15th A Cross Road, Yelahanka Satellite Town, Y...	POINT (77.58393 13.09099)
3	51	Vijnanapura	4	SP Naidu Layout 4th Cross Street, SP Naidu Lay...	POINT (77.66957 13.00606)
4	53	Basavanapura	5	Medahalli Kadugodi Road, Bharathi Nagar, Krish...	POINT (77.71546 13.01685)

Hourly Data:

```
hourly_data1.describe()
```

	sourceid	dstid	hod	mean_travel_time	standard_deviation_travel_time	geometric_mean_travel_time
count	854482.000000	854482.000000	854482.000000	854482.000000	854482.000000	854482.000000
mean	100.098070	100.369252	11.947917	2156.821809	611.681541	2072.412810
std	56.860746	57.365489	6.796431	1159.185407	316.081851	1134.975288
min	1.000000	1.000000	0.000000	29.170000	40.140000	12.390000
25%	51.000000	50.000000	6.000000	1278.090000	397.200000	1214.300000
50%	100.000000	101.000000	12.000000	1987.550000	544.570000	1911.455000
75%	149.000000	150.000000	18.000000	2898.630000	747.050000	2803.967500
max	198.000000	198.000000	23.000000	10873.820000	4728.510000	10672.840000

geometric_standard_deviation_travel_time	mean_travel_time_mins	standard_deviation_travel_time_mins	geometric_mean_travel_time_mins
854482.000000	854482.000000	854482.000000	854482.000000
1.369011	35.947030	10.194692	34.540213
0.275560	19.319757	5.268031	18.916255
1.030000	0.486167	0.669000	0.206500
1.240000	21.301500	6.620000	20.238333
1.300000	33.125833	9.076167	31.857583
1.390000	48.310500	12.450833	46.732792
15.450000	181.230333	78.808500	177.880667

➢ Derived fields:
  ○ mean_travel_time_mins from mean_travel_time
  ○ Standard_deviation_travel_time_mins from Standard_deviation_travel_time
  ○ Geometric_mean_travel_time_mins from Geometric_mean_travel_time
➢ Hourly data gives us time taken to travel between 2 points for every hour of the day
➢ Mean travel time can be 0.5 mins when distance between 2 points are minimum
➢ Mean travel time can be 181.2 mins (3 hrs) when distance between 2 points are minimum

Weekly Data:

```
Weekly_data.describe()
```

	sourceid	dstid	dow	mean_travel_time	standard_deviation_travel_time	geometric_mean_travel_time
count	267662.000000	267662.000000	267662.000000	267662.000000	267662.000000	267662.000000
mean	99.726360	99.768663	4.003684	2348.104783	794.949807	2215.896753
std	57.146814	57.339087	2.000577	1080.959491	332.492531	1037.938717
min	1.000000	1.000000	1.000000	66.760000	53.560000	23.180000
25%	50.000000	50.000000	2.000000	1552.342500	552.510000	1456.962500
50%	100.000000	100.000000	4.000000	2309.110000	748.205000	2186.275000
75%	149.000000	150.000000	6.000000	3091.250000	990.905000	2932.885000
max	198.000000	198.000000	7.000000	7050.330000	3958.520000	6806.120000

geometric_standard_deviation_travel_time	mean_travel_time_mins	standard_deviation_travel_time_mins	geometric_mean_travel_time_mins
267662.000000	267662.000000	267662.000000	267662.000000
1.456126	39.135080	13.249163	36.931613
0.257603	18.015992	5.541542	17.298979
1.060000	1.112667	0.892667	0.386333
1.340000	25.872375	9.208500	24.282708
1.400000	38.485167	12.470083	36.437917
1.480000	51.520833	16.515083	48.881417
8.860000	117.505500	65.975333	113.435333

➢ Derived fields:
  o mean_travel_time_mins from mean_travel_time
  o Standard_deviation_travel_time_mins from Standard_deviation_travel_time
  o Geometric_mean_travel_time_mins from Geometric_mean_travel_time
➢ Weekly data gives us time taken to travel between 2 points on every day of the week
➢ Mean travel time can be minimum 1.1 mins when distance between 2 points are minimum
➢ Mean travel time can be maximum 1.1 mins when distance between 2 points are minimum

Let's dive into actual data with an example:

Choose your travel points:

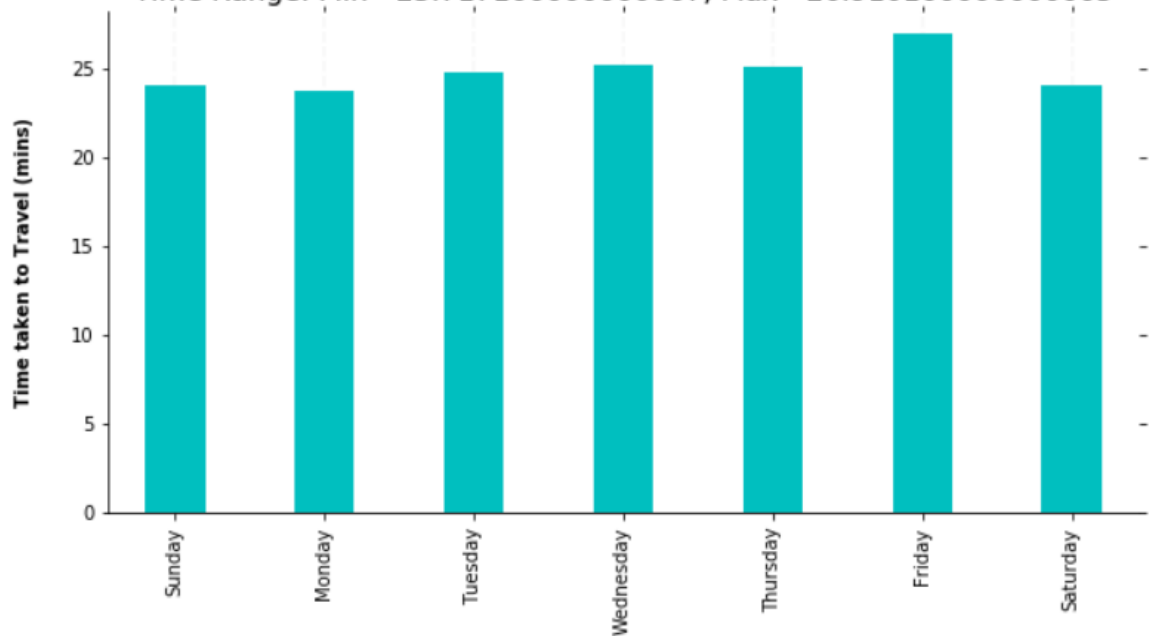➢ Enter the source id: 114

➢ Enter the destination id: 47

Hourly data - describe()

	sourceid	dstid	dow	geometric_standard_deviation_travel_time	mean_travel_time_mins	standard_deviation_travel_time_mins	geometric_mean_travel_time_mins
count	7.0	7.0	7.000000	7.00000	7.000000	7.000000	7.000000
mean	114.0	47.0	4.000000	1.33000	24.821595	7.375429	23.824071
std	0.0	0.0	2.160247	0.01633	1.076759	0.578526	0.999837
min	114.0	47.0	1.000000	1.30000	23.717167	6.387833	22.908833
25%	114.0	47.0	2.500000	1.32500	24.055583	7.036417	23.047167
50%	114.0	47.0	4.000000	1.33000	24.757833	7.647167	23.744667
75%	114.0	47.0	5.500000	1.34000	25.127417	7.725500	24.129083
max	114.0	47.0	7.000000	1.35000	26.910167	8.069167	25.762500

Average travel times by day:

['0 16th Main Road, CHBS Layout, Stage 2, Vijaya Nagar, Bengaluru']
to
['18, Malleshwaram West, Bengaluru']

Time Range: Min - 23.717166666666667, Max - 26.910166666666665

Weekly data - describe()

	sourceid	dstid	hod	geometric_standard_deviation_travel_time	mean_travel_time_mins	standard_deviation_travel_time_mins	geometric_mean_travel_time_mins
count	24.0	24.0	24.000000	24.000000	24.000000	24.000000	24.000000
mean	114.0	47.0	11.500000	1.278333	22.522924	6.175215	21.825910
std	0.0	0.0	7.071068	0.051471	4.996484	0.777987	5.021097
min	114.0	47.0	0.000000	1.210000	15.472500	4.495833	14.952000
25%	114.0	47.0	5.750000	1.240000	16.834792	5.791625	16.059167
50%	114.0	47.0	11.500000	1.260000	24.193667	6.263500	23.609833
75%	114.0	47.0	17.250000	1.320000	26.082417	6.409167	25.454292
max	114.0	47.0	23.000000	1.370000	31.740167	7.922000	30.874833

Average travel times by period:

['6th Cross Road, Lakshman Nagar, State Bank of India Colony, Bengaluru']
to
['700 Ashwath Katte Road, Yeshwanthpur Suburb II Stage, Peenya, Peenya, Bengaluru']

Time Range: Min - 22.12316666666667, Max - 24.753333333333334
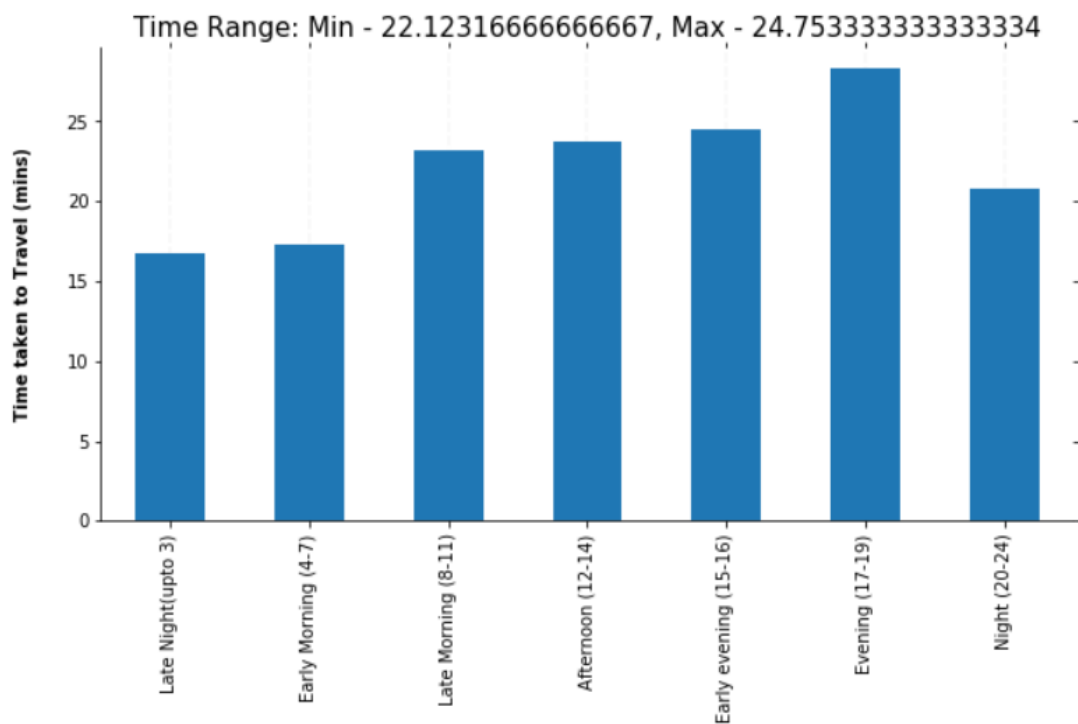
Note: Above graphs can be repeated for geometric mean travel time, standard deviation travel time and geometric standard deviation travel time