Network Analysis of Uber Rides Data

Content

- Dataset
- Data Analysis
  - Number of Rides Based on Type of Purpose
  - Percentage of Rides Based on Type of Category Type
  - Ride Summary - Miles Based on Purpose
  - Number of Rides Based on Months
- Network Analysis
  - Components and Subgraphs
  - Visualization of Uber Traffic Flow Network
  - Analysis of Nodes
  - Network Density
  - Centrality
  - Path Analysis
- Conclusion
- References

1.

Dataset

Features

- Start Date
- End Date
- Start Location
- End Location
- Miles Driven
- Purpose of drive (Meals, Errands, Meetings, Customer Support etc.)
- Category (Business, Personal)

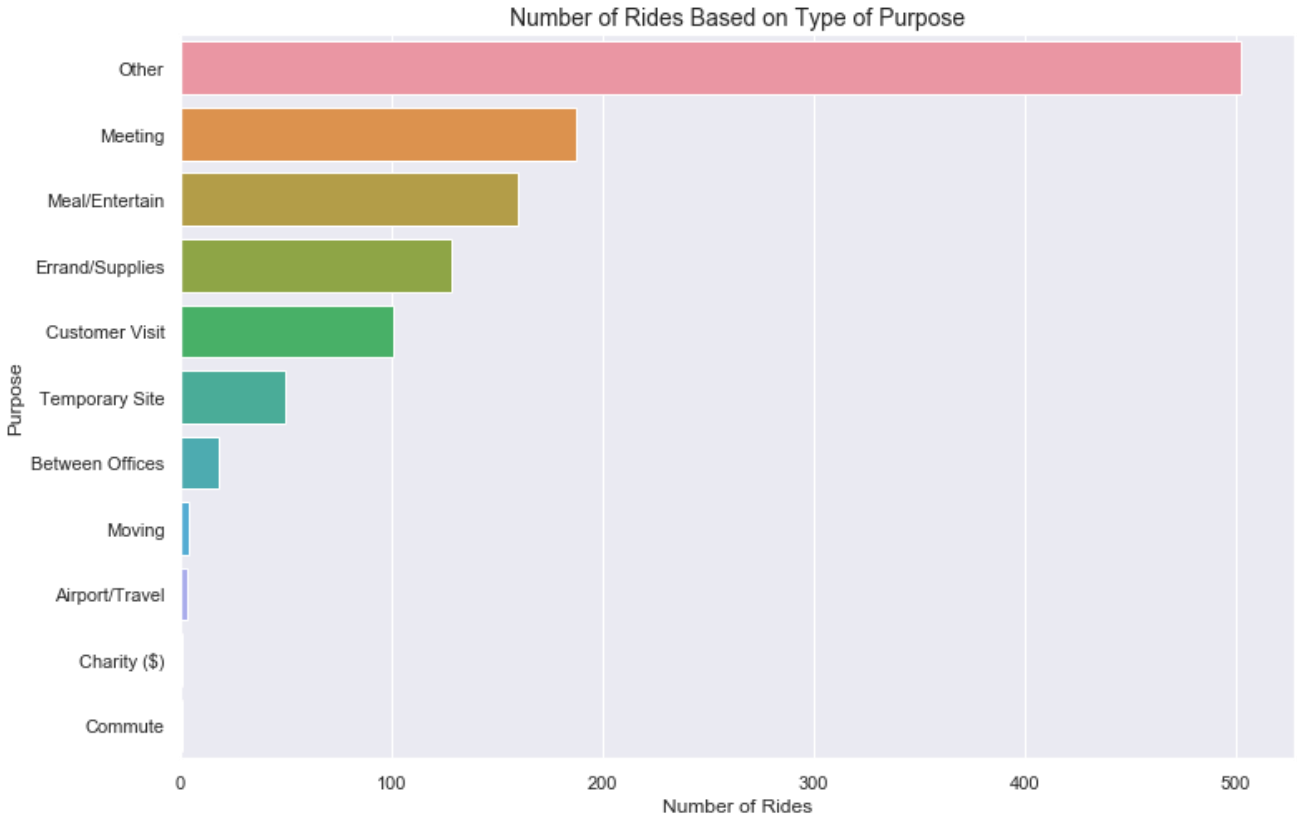
	START_DATE*	END_DATE*	CATEGORY*	START*	STOP*	MILES*	PURPOSE*
0	1/1/2016 21:11	1/1/2016 21:17	Business	Fort Pierce	Fort Pierce	5.1	Meal/Entertain
1	1/2/2016 1:25	1/2/2016 1:37	Business	Fort Pierce	Fort Pierce	5.0	NaN
2	1/2/2016 20:25	1/2/2016 20:38	Business	Fort Pierce	Fort Pierce	4.8	Errand/Supplies
3	1/5/2016 17:31	1/5/2016 17:45	Business	Fort Pierce	Fort Pierce	4.7	Meeting
4	1/6/2016 14:42	1/6/2016 15:49	Business	Fort Pierce	West Palm Beach	63.7	Customer Visit

	count	unique	top	freq
START_DATE*	1155	1154	6/28/2016 23:34	2
END_DATE*	1155	1154	6/28/2016 23:59	2
CATEGORY*	1155	2	Business	1078
START*	1155	177	Cary	201
STOP*	1155	188	Cary	203
PURPOSE*	653	10	Meeting	187

2.

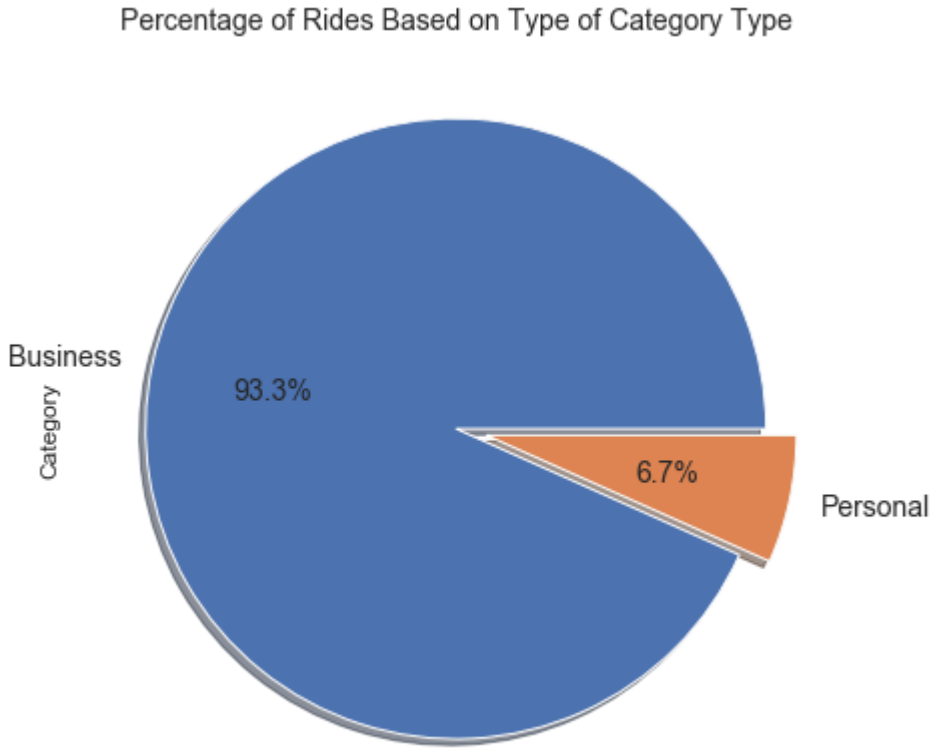
Data Analysis

Number of Rides Based on Type of Purpose

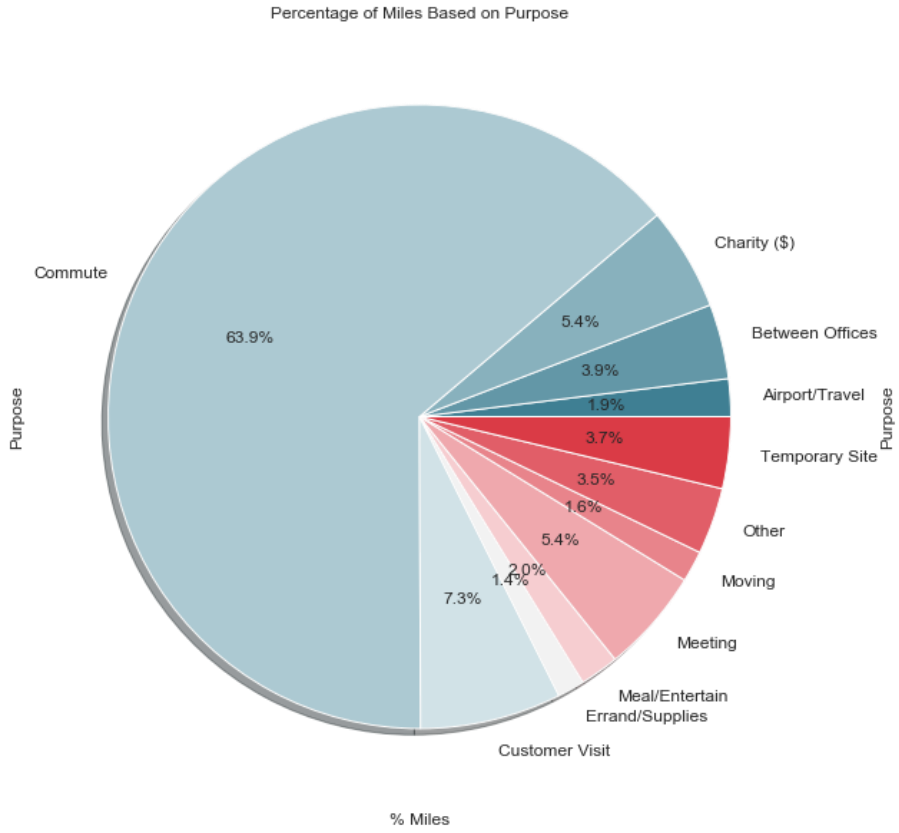


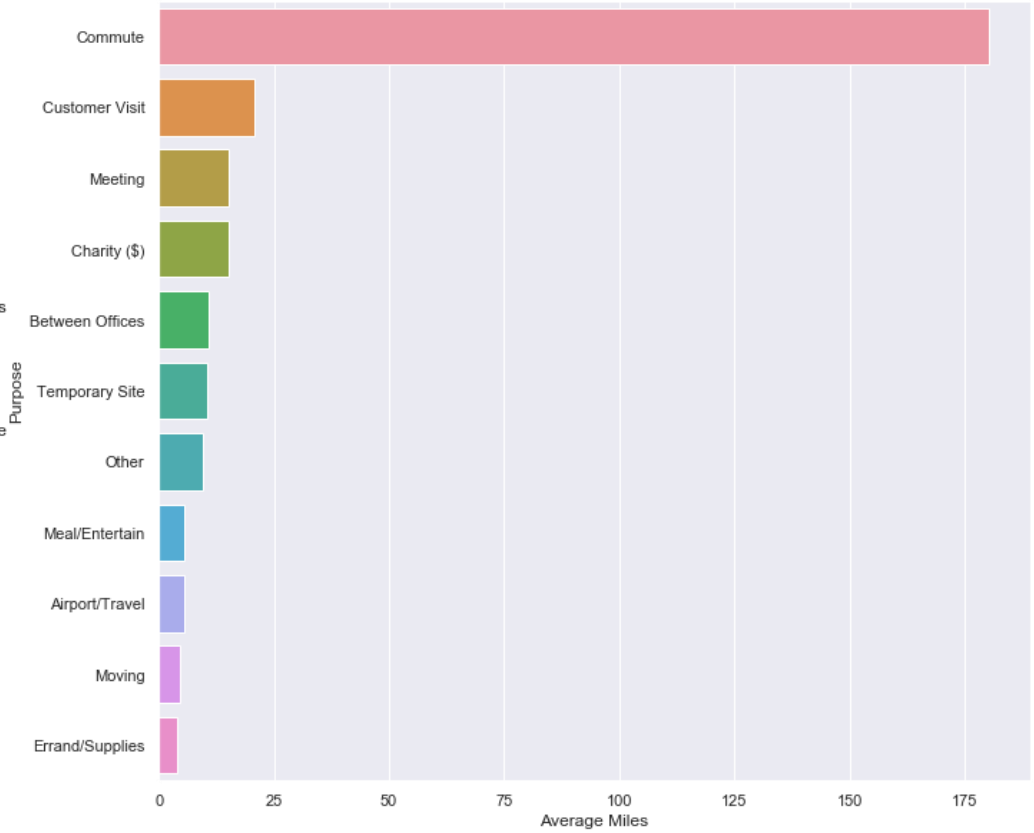
Purpose	
Other	502
Meeting	187
Meal/Entertain	160
Errand/Supplies	128
Customer Visit	101
Temporary Site	50
Between Offices	18
Moving	4
Airport/Travel	3
Commute	1
Charity ($)	1

Percentage of Rides Based on Type of Category Type

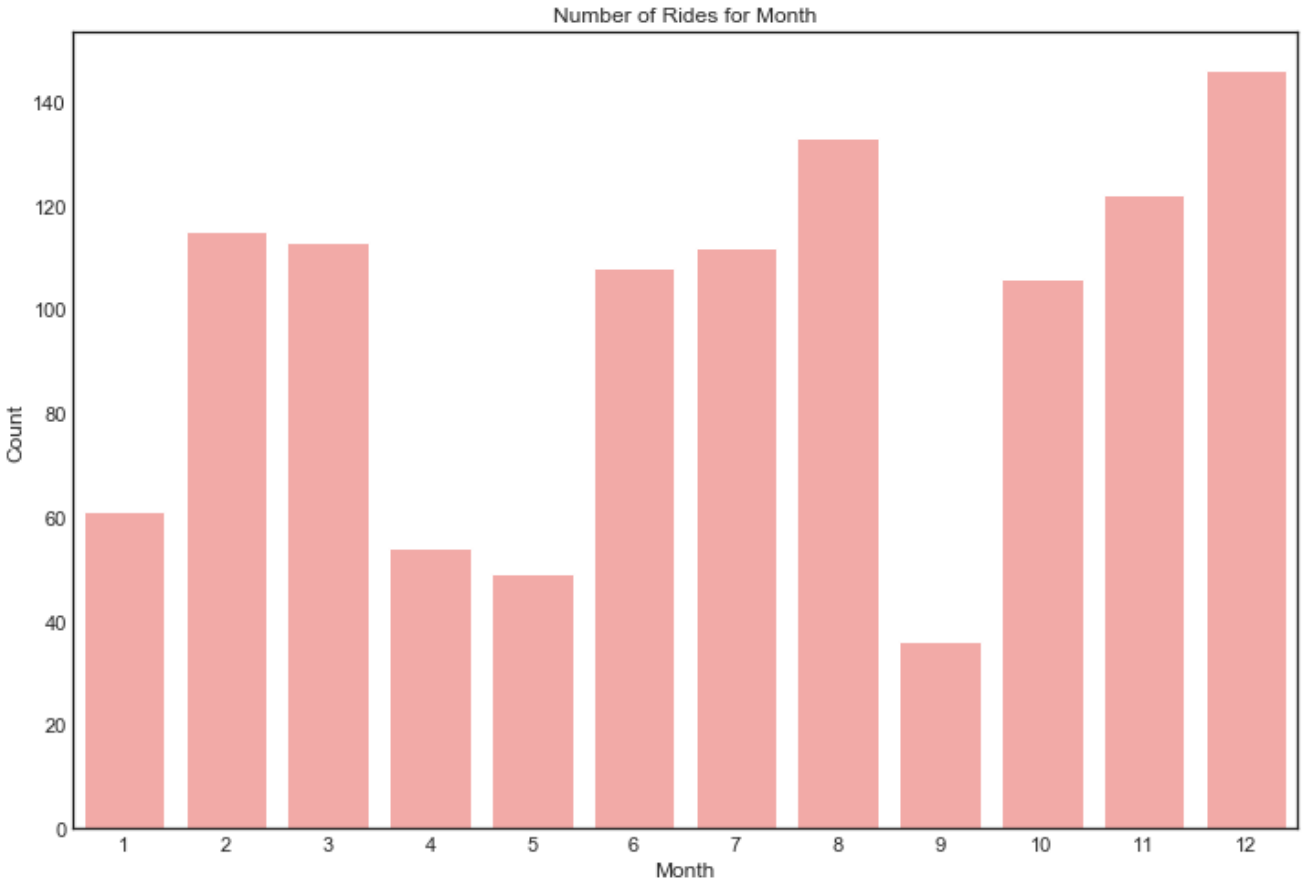


Ride Summary - Miles Based on Purpose





Number of Rides Based on Months



3.
Network Analysis

Graph Creation

- Number of nodes: 209
- Number of edges: 280

The graph has 209 nodes and 280 edges so the size of the network is 209.

On average the degree is 2.67

Nodes & Edges

Nodes:

['Fort Pierce', 'West Palm Beach', 'Palm Beach', 'Cary', 'Morrisville', 'Jamaica', 'New York', 'Queens', 'Elmhurst', 'Midtown', 'East Harlem', 'NoMad', 'Flatiron District', 'Midtown East', 'Hudson Square', 'Lower Manhattan', "Hell's Kitchen", 'Queens County', 'Downtown', 'Gulfton']

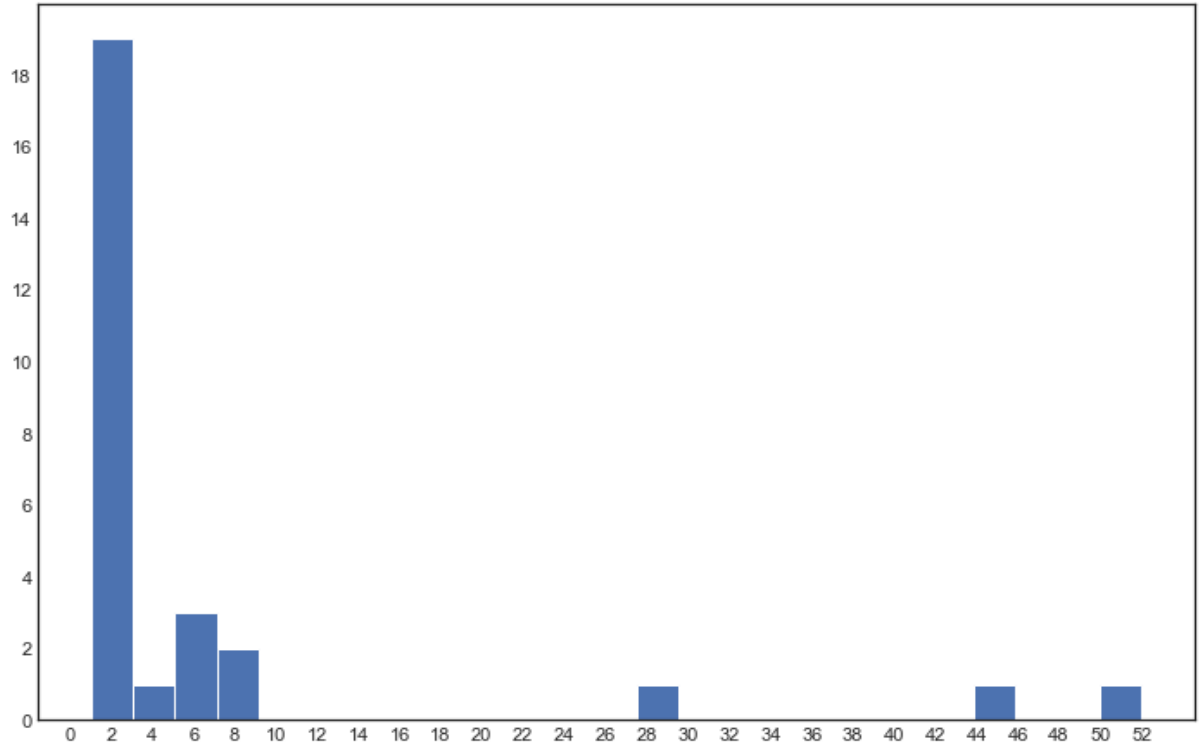
Edges:

[('Fort Pierce', 'Fort Pierce'), ('Fort Pierce', 'West Palm Beach'), ('West Palm Beach', 'West Palm Beach'), ('West Palm Beach', 'Palm Beach'), ('Cary', 'Cary'), ('Cary', 'Morrisville'), ('Cary', 'Durham'), ('Cary', 'Raleigh'), ('Cary', 'Apex'), ('Cary', 'Chapel Hill'), ('Cary', 'Latta'), ('Cary', 'Florence'), ('Cary', 'Holly Springs'), ('Cary', 'Boone'), ('Cary', 'Wake Forest'), ('Cary', 'Eagle Rock'), ('Cary', 'Winston Salem'), ('Cary', 'Mebane'), ('Cary', 'Unknown Location'), ('Cary', 'Wake Co.')]

Components and Subgraphs

- Uber Rides Network contains 209 nodes, but the network is split into over 28 component subgraphs.

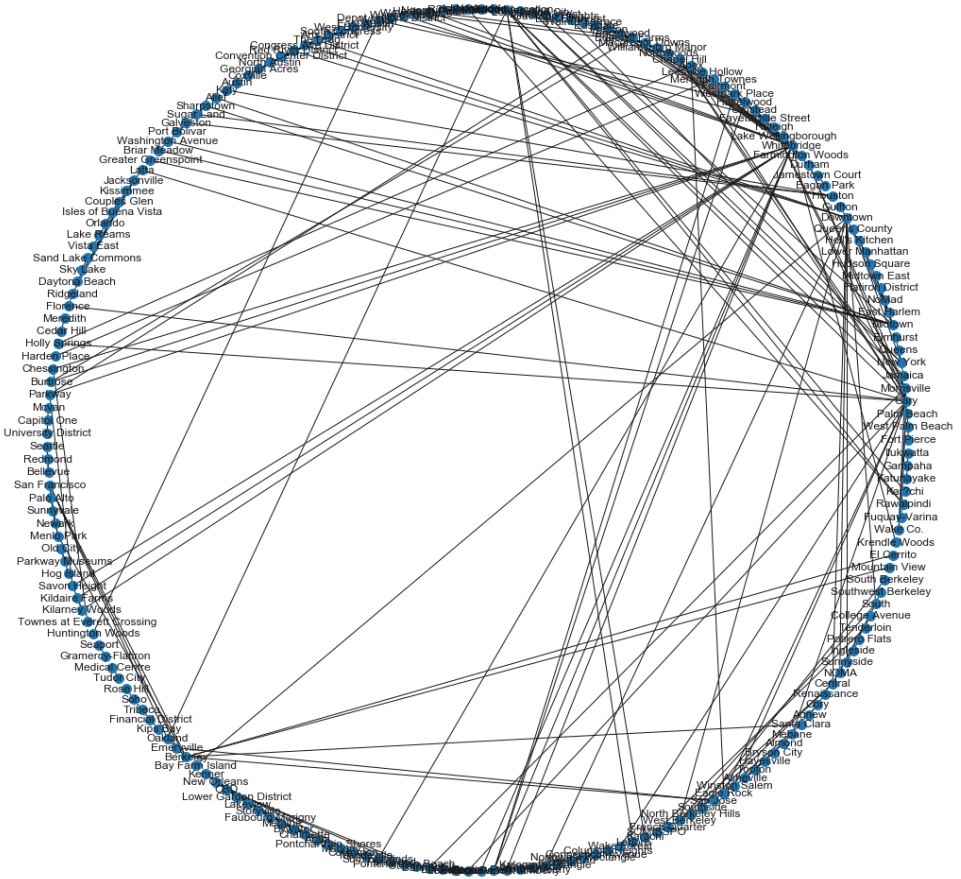
Network Component Distribution



Component	Count
1	3
2	8
3	8
5	1
6	1
7	2
8	1
9	1
44	1
52	1
28	1

- Of 28 components, 3 are of size 1 -these are called "isolates" and should be removed from the network-.
- There are 8 components of size 2 and 3. And the biggest subgraph is contains 52 nodes of a network.

Visualization of Uber Traffic Flow Network

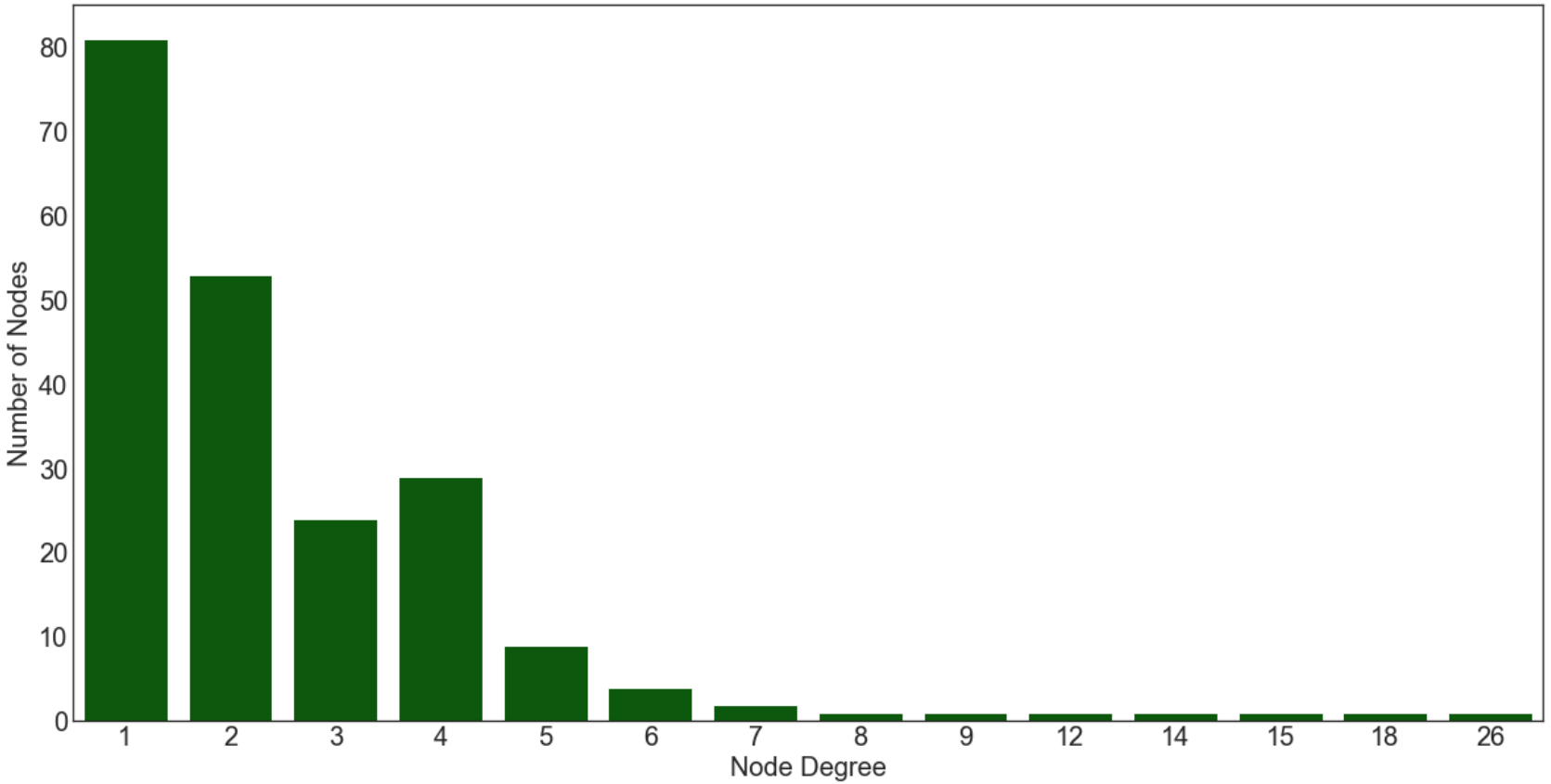


Analysis of Nodes

	Location	Degree
0	Whitebridge	26
1	Cary	18
2	Unknown Location	15
3	Midtown	14
4	Downtown	12
5	Morrisville	9
6	Berkeley	8
8	Central	7
7	The Drag	7
9	New York	6
10	Apex	6
11	Islamabad	6
12	Banner Elk	6
17	Kissimmee	5
21	Bryson City	5
20	West Berkeley	5
19	Lexington Park at Amberly	5
18	San Francisco	5
15	Preston	5
14	Edgehill Farms	5

- Top 20 nodes by degree is shown. The results identifies which location is being visited more frequently.
  - It can be seen that Whitebridge and Cary are the places that are visited mostly.

Distribution of the node degrees



Most of the nodes tend to have a degree of 1 or 2 and 4.

Network Density

Network density: 0.012881854987118146

- The density of our network is approximately 0.0129.
- On a scale of 0 to 1, not a very dense network, which comports with what you can see in the visualization. A 0 would mean that there are no connections at all, and a 1 would indicate that all possible edges are present (a perfectly connected network):
  - This Uber network is on the lower end of that scale, but still far from 0.

Centrality

Degree Centrality

```
[('Whitebridge', 0.125),
 ('Cary', 0.08653846153846154),
 ('Unknown Location', 0.07211538461538462),
 ('Midtown', 0.06730769230769232),
 ('Downtown', 0.057692307692307696),
 ('Morrisville', 0.04326923076923077),
 ('Berkeley', 0.038461538461538464),
 ('The Drag', 0.03365384615384616),
 ('Central', 0.03365384615384616),
 ('New York', 0.028846153846153848)]
```

Eigenvalue Centrality

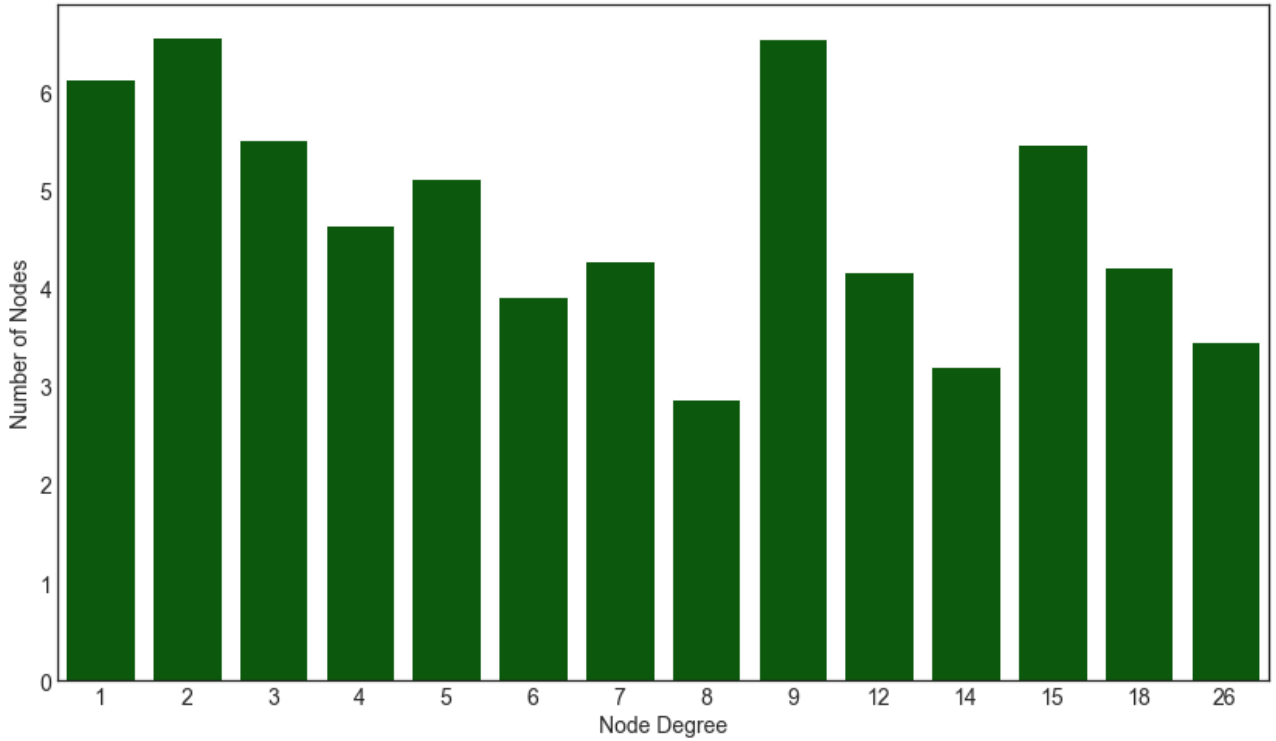
```
[('Whitebridge', 0.6646598425612025),
 ('Preston', 0.2258961119723626),
 ('Lexington Park at Amberly', 0.20712910313780167),
 ('Edgehill Farms', 0.2051712166317349),
 ('Westpark Place', 0.20047189476573732),
 ('Tanglewood', 0.19234876125053213),
 ('Hazelwood', 0.17431819214287014),
 ('Northwoods', 0.17066934400969044),
 ('Parkway', 0.15923630129510694),
 ('Savon Height', 0.1533529382511316)]
```

Katz Centrality

```
[('Whitebridge', 0.2719384341254852),
 ('Cary', 0.1951184803643446),
 ('Unknown Location', 0.17465046022598132),
 ('Midtown', 0.13451560365297247),
 ('Downtown', 0.13322538808045487),
 ('Morrisville', 0.13251870013513256),
 ('Preston', 0.10870972414143357),
 ('Lexington Park at Amberly', 0.105085109926472),
 ('Edgehill Farms', 0.10424685821946111),
 ('Westpark Place', 0.10043500585941184)]
```

Average Degree Connectivity

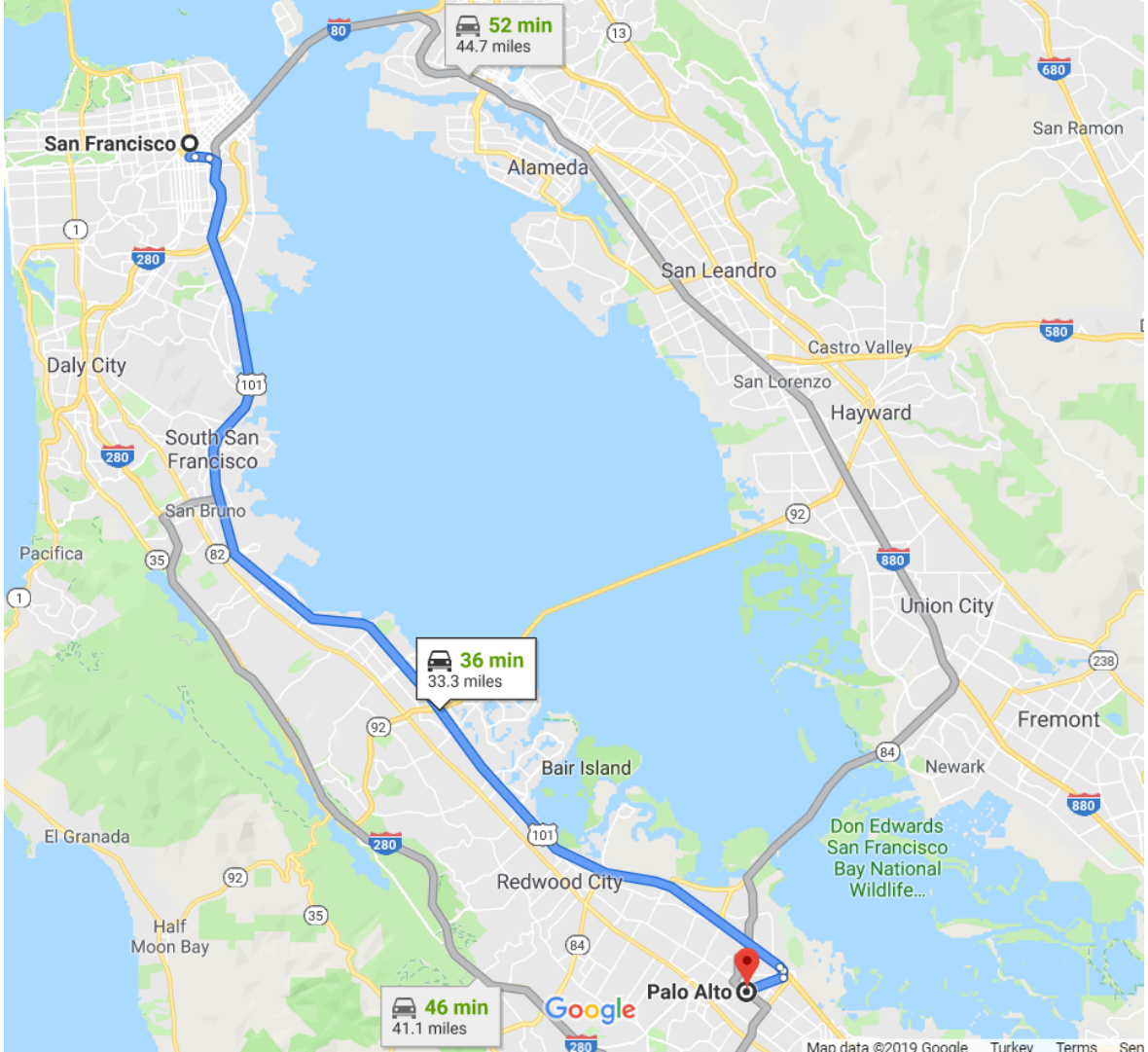
	Avg Degree of Neighbors
1	6.1358
2	6.56604
3	5.52778
4	4.65517
5	5.13333
6	3.91667
7	4.28571
8	2.875
9	6.55556
12	4.16667
14	3.21429
15	5.46667
18	4.22222
26	3.46154



Path Analysis

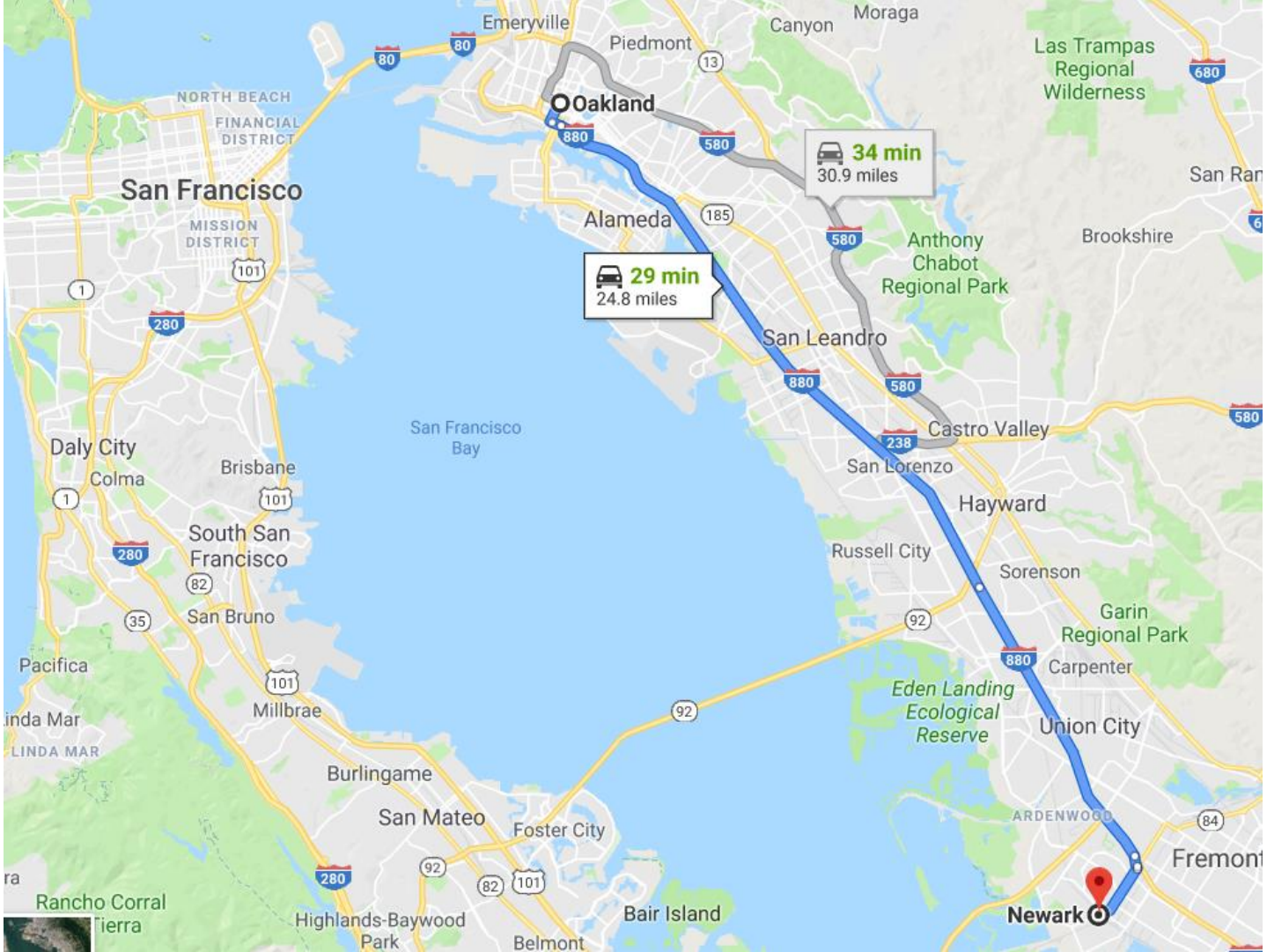
All available paths from San Francisco to Palo Alto:

```
['San Francisco', 'Palo Alto']
['San Francisco', 'Newark', 'Sunnyvale', 'Palo Alto']
['San Francisco', 'Newark', 'Menlo Park', 'Palo Alto']
['San Francisco', 'Emeryville', 'Oakland', 'Berkeley', 'Menlo Park', 'Newark', 'Sunnyvale', 'Palo Alto']
['San Francisco', 'Emeryville', 'Oakland', 'Berkeley', 'Menlo Park', 'Palo Alto']
['San Francisco', 'Emeryville', 'Berkeley', 'Menlo Park', 'Newark', 'Sunnyvale', 'Palo Alto']
['San Francisco', 'Emeryville', 'Berkeley', 'Menlo Park', 'Palo Alto']
['San Francisco', 'Emeryville', 'San Jose', 'Berkeley', 'Menlo Park', 'Newark', 'Sunnyvale', 'Palo Alto']
['San Francisco', 'Emeryville', 'San Jose', 'Berkeley', 'Menlo Park', 'Palo Alto']
['San Francisco', 'Emeryville', 'San Jose', 'Santa Clara', 'Berkeley', 'Menlo Park', 'Newark', 'Sunnyvale', 'Palo Alto']
['San Francisco', 'Emeryville', 'San Jose', 'Santa Clara', 'Berkeley', 'Menlo Park', 'Palo Alto']
['San Francisco', 'Oakland', 'Emeryville', 'Berkeley', 'Menlo Park', 'Newark', 'Sunnyvale', 'Palo Alto']
['San Francisco', 'Oakland', 'Emeryville', 'Berkeley', 'Menlo Park', 'Palo Alto']
['San Francisco', 'Oakland', 'Emeryville', 'San Jose', 'Berkeley', 'Menlo Park', 'Newark', 'Sunnyvale', 'Palo Alto']
['San Francisco', 'Oakland', 'Emeryville', 'San Jose', 'Berkeley', 'Menlo Park', 'Palo Alto']
['San Francisco', 'Oakland', 'Emeryville', 'San Jose', 'Santa Clara', 'Berkeley', 'Menlo Park', 'Newark', 'Sunnyvale', 'Palo Al
to']
['San Francisco', 'Oakland', 'Emeryville', 'San Jose', 'Santa Clara', 'Berkeley', 'Menlo Park', 'Palo Alto']
['San Francisco', 'Oakland', 'Berkeley', 'Menlo Park', 'Newark', 'Sunnyvale', 'Palo Alto']
['San Francisco', 'Oakland', 'Berkeley', 'Menlo Park', 'Palo Alto']
['San Francisco', 'Berkeley', 'Menlo Park', 'Newark', 'Sunnyvale', 'Palo Alto']
['San Francisco', 'Berkeley', 'Menlo Park', 'Palo Alto']
```


52 min
44.7 miles
San Francisco
Alameda
San Ramon
San Leandro
Castro Valley
San Lorenzo
Hayward
Daly City
South San Francisco
San Bruno
Pacifica
Union City
36 min
33.3 miles
Fremont
Bair Island
Newark
El Granada
Don Edwards San Francisco Bay National Wildlife...
Redwood City
Half Moon Bay
46 min
41.1 miles
Palo Alto
Google
Map data ©2019 Google
Turkey
Terms

All available paths from Oakland to Newark:

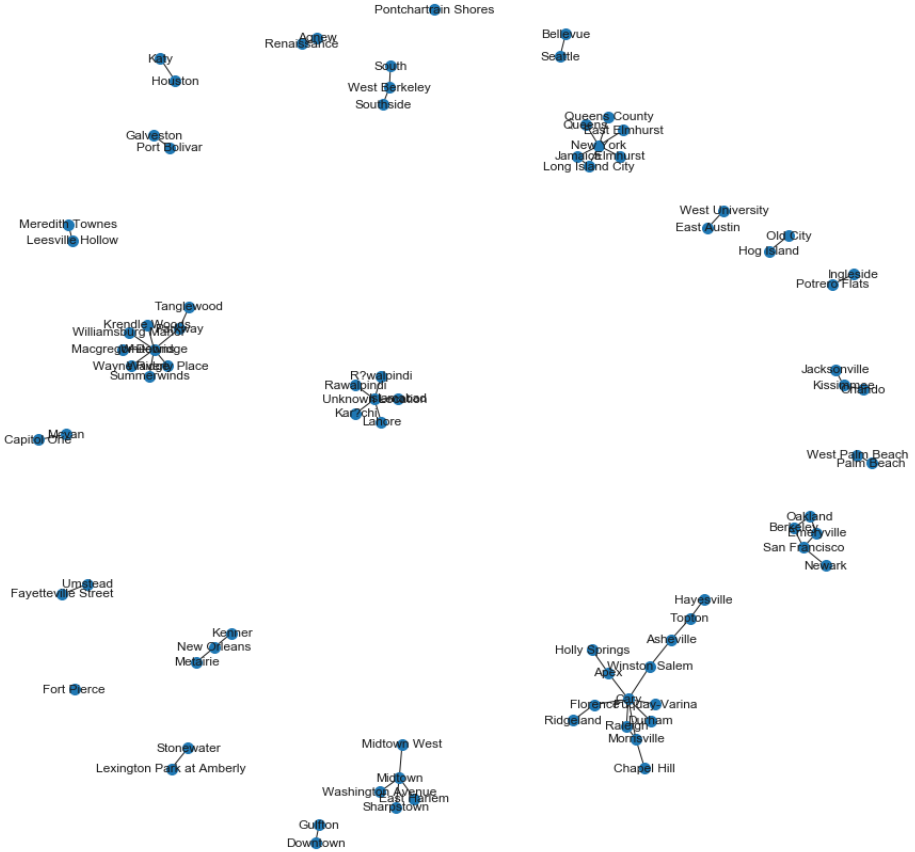
```
['Oakland', 'Emeryville', 'Berkeley', 'San Francisco', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'Emeryville', 'Berkeley', 'San Francisco', 'Palo Alto', 'Menlo Park', 'Newark']
['Oakland', 'Emeryville', 'Berkeley', 'San Francisco', 'Newark']
['Oakland', 'Emeryville', 'Berkeley', 'Menlo Park', 'Newark']
['Oakland', 'Emeryville', 'Berkeley', 'Menlo Park', 'Palo Alto', 'San Francisco', 'Newark']
['Oakland', 'Emeryville', 'Berkeley', 'Menlo Park', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'Emeryville', 'San Francisco', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'Emeryville', 'San Francisco', 'Palo Alto', 'Menlo Park', 'Newark']
['Oakland', 'Emeryville', 'San Francisco', 'Newark']
['Oakland', 'Emeryville', 'San Francisco', 'Berkeley', 'Menlo Park', 'Newark']
['Oakland', 'Emeryville', 'San Francisco', 'Berkeley', 'Menlo Park', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'Emeryville', 'San Jose', 'Berkeley', 'San Francisco', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'Emeryville', 'San Jose', 'Berkeley', 'San Francisco', 'Palo Alto', 'Menlo Park', 'Newark']
['Oakland', 'Emeryville', 'San Jose', 'Berkeley', 'San Francisco', 'Newark']
['Oakland', 'Emeryville', 'San Jose', 'Berkeley', 'Menlo Park', 'Newark']
['Oakland', 'Emeryville', 'San Jose', 'Berkeley', 'Menlo Park', 'Palo Alto', 'San Francisco', 'Newark']
['Oakland', 'Emeryville', 'San Jose', 'Berkeley', 'Menlo Park', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'Emeryville', 'San Jose', 'Santa Clara', 'Berkeley', 'San Francisco', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'Emeryville', 'San Jose', 'Santa Clara', 'Berkeley', 'San Francisco', 'Palo Alto', 'Menlo Park', 'Newark']
['Oakland', 'Emeryville', 'San Jose', 'Santa Clara', 'Berkeley', 'San Francisco', 'Newark']
['Oakland', 'Emeryville', 'San Jose', 'Santa Clara', 'Berkeley', 'Menlo Park', 'Newark']
['Oakland', 'Emeryville', 'San Jose', 'Santa Clara', 'Berkeley', 'Menlo Park', 'Palo Alto', 'San Francisco', 'Newark']
['Oakland', 'Emeryville', 'San Jose', 'Santa Clara', 'Berkeley', 'Menlo Park', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'Berkeley', 'Emeryville', 'San Francisco', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'Berkeley', 'Emeryville', 'San Francisco', 'Palo Alto', 'Menlo Park', 'Newark']
['Oakland', 'Berkeley', 'Emeryville', 'San Francisco', 'Newark']
['Oakland', 'Berkeley', 'San Francisco', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'Berkeley', 'San Francisco', 'Palo Alto', 'Menlo Park', 'Newark']
['Oakland', 'Berkeley', 'San Francisco', 'Newark']
['Oakland', 'Berkeley', 'San Jose', 'Emeryville', 'San Francisco', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'Berkeley', 'San Jose', 'Emeryville', 'San Francisco', 'Palo Alto', 'Menlo Park', 'Newark']
['Oakland', 'Berkeley', 'San Jose', 'Emeryville', 'San Francisco', 'Newark']
['Oakland', 'Berkeley', 'Santa Clara', 'San Jose', 'Emeryville', 'San Francisco', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'Berkeley', 'Santa Clara', 'San Jose', 'Emeryville', 'San Francisco', 'Palo Alto', 'Menlo Park', 'Newark']
['Oakland', 'Berkeley', 'Santa Clara', 'San Jose', 'Emeryville', 'San Francisco', 'Newark']
['Oakland', 'Berkeley', 'Menlo Park', 'Newark']
['Oakland', 'Berkeley', 'Menlo Park', 'Palo Alto', 'San Francisco', 'Newark']
['Oakland', 'Berkeley', 'Menlo Park', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'San Francisco', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'San Francisco', 'Palo Alto', 'Menlo Park', 'Newark']
['Oakland', 'San Francisco', 'Newark']
['Oakland', 'San Francisco', 'Emeryville', 'Berkeley', 'Menlo Park', 'Newark']
['Oakland', 'San Francisco', 'Emeryville', 'Berkeley', 'Menlo Park', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'San Francisco', 'Emeryville', 'San Jose', 'Berkeley', 'Menlo Park', 'Newark']
['Oakland', 'San Francisco', 'Emeryville', 'San Jose', 'Berkeley', 'Menlo Park', 'Palo Alto', 'Sunnyvale', 'Newark']
['Oakland', 'San Francisco', 'Emeryville', 'San Jose', 'Santa Clara', 'Berkeley', 'Menlo Park', 'Newark']
['Oakland', 'San Francisco', 'Emeryville', 'San Jose', 'Santa Clara', 'Berkeley', 'Menlo Park', 'Palo Alto', 'Sunnyvale', 'Newa
rk']
['Oakland', 'San Francisco', 'Berkeley', 'Menlo Park', 'Newark']
['Oakland', 'San Francisco', 'Berkeley', 'Menlo Park', 'Palo Alto', 'Sunnyvale', 'Newark']
```


Emeryville
Piedmont
Canyon
Moraga
Las Trampas Regional Wilderness
NORTH BEACH
FINANCIAL DISTRICT
Oakland
34 min
30.9 miles
San Francisco
MISSION DISTRICT
Alameda
Anthony Chabot Regional Park
Brookshire
29 min
24.8 miles
San Leandro
San Francisco Bay
Daly City
Colma
Brisbane
Castro Valley
San Lorenzo
Hayward
South San Francisco
Russell City
Sorenson
San Bruno
Garin Regional Park
Pacifica
Carpenter
Millbrae
Eden Landing Ecological Reserve
Union City
LINDA MAR
Burlingame
San Mateo
Foster City
ARDENWOOD
Fremont
Rancho Corral
Highlands-Baywood Park
Bair Island
Belmont
Newark

dijkstra_path(G, source='Oakland', target='Newark')

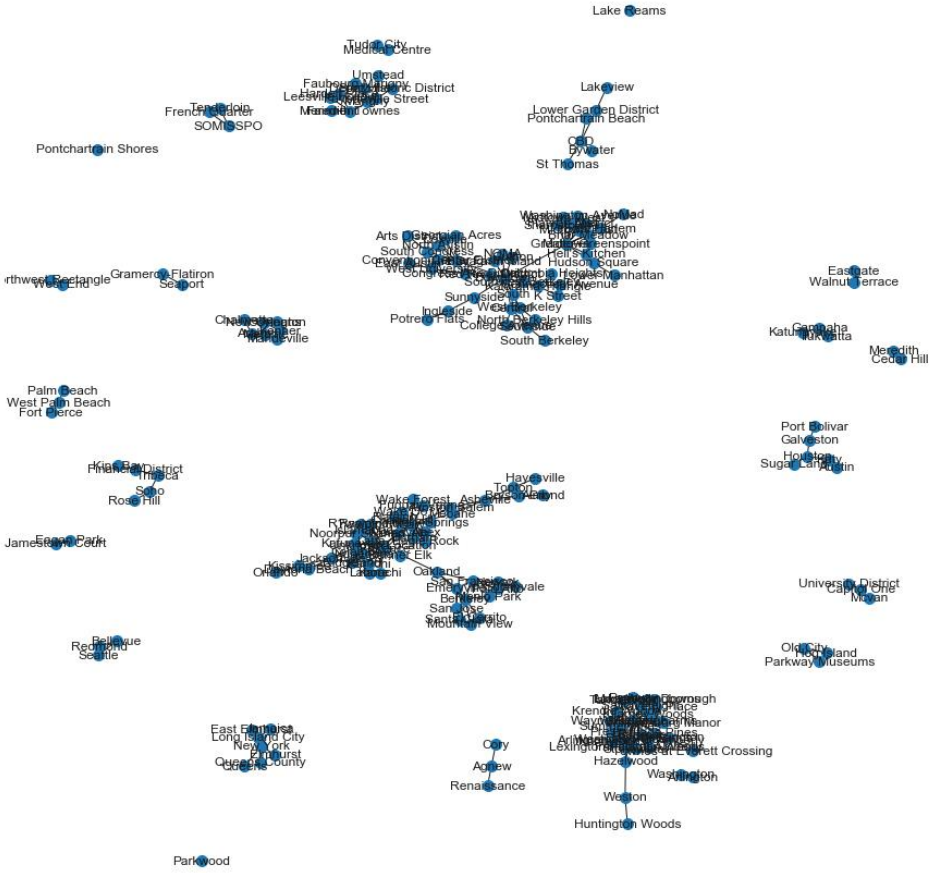
➢ ['Oakland', 'San Francisco', 'Newark']

Visualization of Network of Rides Based on Meeting Purpose



	Location	Degree
0	Cary	9
2	Unknown Location	7
1	Whitebridge	7
3	New York	6
4	Midtown	4
5	Kissimmee	4
10	Lahore	3
11	Kar?chi	3
9	San Francisco	3
8	R?walpindi	3
7	Islamabad	3
6	Morrisville	3
20	Emeryville	2
26	Berkeley	2
25	Topton	2
24	Asheville	2
23	Winston Salem	2
22	New Orleans	2
21	Pontchartrain Shores	2
27	West Berkeley	2

Visualization of Network of Rides Based on Business Category Type



	Location	Degree
0	Whitebridge	26
1	Cary	17
2	Unknown Location	15
3	Midtown	14
4	Downtown	12
5	Morrisville	9
6	Berkeley	8
7	The Drag	7
8	Central	7
9	New York	6
10	Apex	6
11	Islamabad	6
16	Lexington Park at Amberly	5
18	Bryson City	5
17	West Berkeley	5
15	San Francisco	5
14	Kissimmee	5
13	Colombo	5
12	Houston	5
32	Kenner	4

4.
Conclusion

- The highest number of rides mostly made for the purpose of meeting.
- The highest miles are taken for the commuting purpose.
- The most number of rides made on the August and the December.
- The graph has 209 nodes and 280 edges so the size of the network is 209. The average degree size is 2.67.
- Start and End locations are used as nodes of the network and for edge attributes the remaining features (StartDate, EndDate, Category, Miles, Purpose) are used.
- The network is split into over 28 component subgraphs.
- Whitebridge and Cary are the places that are visited mostly based on node degrees.
- The density of our network is approximately 0.0129. It is not a very dense network.

References

- http://snap.stanford.edu/class/cs224w-2017/projects/cs224w-11-final.pdf
- https://www.kaggle.com/zusmani/uberdrives