

CS480 Ridesharing Project

Algorithm

- filter by passengers (single riders and double riders)
 - 1 rider + 1 rider
 - 1 rider + 2 rider
- filter by the time window (difference between each trip from one another)
- group by location (PULocationID, DULocationID)
- social preferences (do they want to talk, do they want to share a ride, etc. but that info is not available)

```
In [36]: import pandas as pd  
import numpy as np
```

1 Setup

1.1 Load Yellow Trip from December 2016 dataset

```
In [25]: # Load Yellow Trip from December 2016 dataset
yellow_taxi = pd.read_csv('yellow_tripdata_2019-12.csv', dtype={'stor
e_and_fwd_flag': object})
yellow_taxi
```

Out[25]:

	VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance
0	1.0	2019-12-01 00:26:58	2019-12-01 00:41:45	1.0	4.2
1	1.0	2019-12-01 00:12:08	2019-12-01 00:12:14	1.0	0.0
2	1.0	2019-12-01 00:25:53	2019-12-01 00:26:04	1.0	0.0
3	1.0	2019-12-01 00:12:03	2019-12-01 00:33:19	2.0	9.4
4	1.0	2019-12-01 00:05:27	2019-12-01 00:16:32	2.0	1.6
...
6896312	NaN	2019-12-31 00:07:00	2019-12-31 00:46:00	NaN	12.7
6896313	NaN	2019-12-31 00:20:00	2019-12-31 00:47:00	NaN	18.5
6896314	NaN	2019-12-31 00:50:00	2019-12-31 01:21:00	NaN	13.1
6896315	NaN	2019-12-31 00:38:19	2019-12-31 01:19:37	NaN	14.5
6896316	NaN	2019-12-31 00:21:00	2019-12-31 00:56:00	NaN	-17.1

6896317 rows × 18 columns

1.2 Total rides traveled in yellow taxis in December 2016

```
In [37]: total_rides = len(yellow_taxi.index)
print("There were {0} total rides in yellow taxis in December 2016.".f
ormat(total_rides))
```

There were 6896317 total rides in yellow taxis in December 2016.

1.3 Total Distance traveled in all the trips

```
In [28]: total_distance = yellow_taxi['trip_distance'].sum()
print("Total distance is {0} miles in yellow taxis in December 2016.".
format(total_distance))
```

Total distance is 20505652.880000003 miles in yellow taxis in December 2016.

2 Filter Passengers

2.1.a Gather the single riders

```
In [47]: # the rides in yellow taxis that can be combined based on the number of
# passengers
# and the drop off is within the 50 - 100 drop off region
single_riders = yellow_taxi.loc[yellow_taxi['passenger_count'] == 1.0]
single_riders
```

Out[47]:

	VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance
0	1.0	2019-12-01 00:26:58	2019-12-01 00:41:45	1.0	4.2
1	1.0	2019-12-01 00:12:08	2019-12-01 00:12:14	1.0	0.0
2	1.0	2019-12-01 00:25:53	2019-12-01 00:26:04	1.0	0.0
9	1.0	2019-12-01 00:19:48	2019-12-01 00:24:18	1.0	0.9
11	2.0	2019-11-30 23:54:45	2019-12-01 00:09:57	1.0	2.1
...
6845293	2.0	2019-12-31 23:20:52	2019-12-31 23:25:16	1.0	0.7
6845294	2.0	2019-12-31 23:09:42	2019-12-31 23:22:39	1.0	1.4
6845295	2.0	2019-12-31 23:46:57	2019-12-31 23:50:54	1.0	0.9
6845296	2.0	2019-12-31 23:22:03	2019-12-31 23:29:06	1.0	3.3
6845297	2.0	2019-12-31 23:33:47	2019-12-31 23:40:10	1.0	1.5

4783392 rows × 6 columns

2.1.b Total rides traveled in yellow taxis with a single rider in December 2016

```
In [44]: num_single_riders = len(single_riders.index)
print("There were {0} total rides with single riders in yellow taxis in December 2016.".format(num_single_riders))
```

There were 4783392 total rides with single riders in yellow taxis in December 2016.

2.1.c Total distance traveled in yellow taxis with single riders in December 2016

```
In [40]: total_distance_single_riders = single_riders['trip_distance'].sum()
print("Total distance is {0} miles in yellow taxis with single riders in December 2016.".format(total_distance_single_riders))
```

Total distance is 13797069.180000003 miles in yellow taxis with single riders in December 2016.

2.2.a Gather the double riders

```
In [45]: # the rides in yellow taxis that can be combined based on the number of passengers
double_riders = yellow_taxis.loc[yellow_taxis['passenger_count'] == 2.0]
double_riders
```

Out[45]:

	VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance
3	1.0	2019-12-01 00:12:03	2019-12-01 00:33:19	2.0	9.4
4	1.0	2019-12-01 00:05:27	2019-12-01 00:16:32	2.0	1.6
5	1.0	2019-12-01 00:58:51	2019-12-01 01:08:37	2.0	1.0
19	1.0	2019-12-01 00:43:09	2019-12-01 01:11:07	2.0	11.3
31	1.0	2019-12-01 00:01:45	2019-12-01 00:13:26	2.0	4.7
...
6845280	1.0	2019-12-31 23:05:12	2019-12-31 23:13:06	2.0	1.8
6845281	1.0	2019-12-31 23:26:03	2019-12-31 23:47:35	2.0	5.9
6845282	2.0	2019-12-31 23:15:03	2019-12-31 23:25:16	2.0	2.6
6845284	1.0	2019-12-31 23:07:40	2019-12-31 23:20:56	2.0	3.4
6845288	1.0	2019-12-31 23:08:32	2019-12-31 23:24:59	2.0	2.1

1075747 rows × 6 columns

2.2.b Total rides traveled in yellow taxis with double riders in December 2016

```
In [49]: num_double_riders = len(double_riders.index)
print("There were {0} total rides with single riders in yellow taxis in December 2016.".format(num_double_riders))
```

There were 1075747 total rides with single riders in yellow taxis in December 2016.

2.2.c Total rides traveled in yellow taxis with double riders in December 2016

```
In [50]: total_distance_double_riders = double_riders['trip_distance'].sum()  
print("Total distance is {0} miles in yellow taxis with single riders  
in December 2016.".format(total_distance_double_riders))
```

Total distance is 3397516.2000000007 miles in yellow taxis with single riders in December 2016.

3 Filter By Pick Up and Drop Off Location

3.1.a Filtering the single riders by the pickup id from 0-50

```
In [72]: single_0_50 = single_riders.loc[yellow_single_riders['PULocationID'] >  
0]  
single_0_50 = single_0_50.loc[single_0_50['PULocationID'] < 50]  
single_0_50
```

Out[72]:

	VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance
13	2.0	2019-12-01 00:43:02	2019-12-01 01:11:18	1.0	13.0
21	2.0	2019-12-01 00:10:31	2019-12-01 00:27:13	1.0	2.2
32	1.0	2019-12-01 00:20:28	2019-12-01 00:25:32	1.0	0.6
46	1.0	2019-12-01 00:37:48	2019-12-01 00:56:50	1.0	5.2
50	2.0	2019-12-01 00:22:16	2019-12-01 00:36:14	1.0	2.7
...
6845210	2.0	2019-12-31 23:09:06	2019-12-31 23:41:18	1.0	17.4
6845211	2.0	2019-12-31 23:49:05	2019-12-31 23:54:20	1.0	1.1
6845215	2.0	2019-12-31 23:25:02	2019-12-31 23:43:14	1.0	1.2
6845250	2.0	2019-12-31 23:23:19	2020-01-01 00:03:22	1.0	1.9
6845263	1.0	2019-12-31 23:17:19	2019-12-31 23:32:11	1.0	3.3

350418 rows × 6 columns

3.1.b Total single riders being picked up in locationIDS from 0-50

```
In [79]: num_single_riders_0_50 = len(single_0_50.index)
print("There were {0} rides with single riders with pick ups within 0-50 LocationIDs in December 2016.".format(num_single_riders_0_50))
```

There were 55316 rides with single riders with pick ups within 0-50 LocationIDs in December 2016.

3.1.c Filtering by the dropoff id from 0-50

```
In [80]: #from the riders being picked up in 0-10, who is being dropped off in 0-10
single_0_50 = single_0_50.loc[single_0_50['DOLocationID'] > 0]
single_0_50 = single_0_50.loc[single_0_50['DOLocationID'] < 50]
single_0_50
```

Out[80]:

	VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance
61	2.0	2019-12-01 00:39:00	2019-12-01 00:57:35	1.0	4.9
269	2.0	2019-12-01 00:36:17	2019-12-01 01:01:13	1.0	7.2
349	1.0	2019-12-01 00:03:43	2019-12-01 00:05:24	1.0	0.0
373	2.0	2019-12-01 00:22:11	2019-12-01 00:26:33	1.0	0.7
531	1.0	2019-12-01 00:52:22	2019-12-01 00:56:15	1.0	0.8
...
6844724	2.0	2019-12-31 23:00:28	2019-12-31 23:07:28	1.0	0.9
6844725	2.0	2019-12-31 23:12:18	2019-12-31 23:18:00	1.0	1.1
6844754	1.0	2019-12-31 23:14:04	2019-12-31 23:23:12	1.0	1.6
6844755	1.0	2019-12-31 23:35:07	2019-12-31 23:40:02	1.0	0.8
6845195	2.0	2019-12-31 23:20:56	2019-12-31 23:29:52	1.0	1.2

55316 rows × 6 columns

3.1.d Total single riders being picked up AND dropped off in locationIDs from 0-50

```
In [81]: num_single_riders_0_50 = len(single_0_50.index)
print("There were {0} rides with single riders with pickups and drop offs in locationIDs within 0-50 in December 2016.".format(num_single_riders_0_50))
```

There were 55316 rides with single riders with pickups and drop offs in locationIDs within 0-50 in December 2016.

3.2.a Filtering the double riders by the pickup id from 0-50

```
In [82]: double_riders_0_50 = double_riders.loc[double_riders['PULocationID'] > 0]
double_riders_0_50 = double_riders_0_50.loc[double_riders_0_50['PULocationID'] < 50]
double_riders_0_50
```

Out[82]:

	VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance
72	1.0	2019-12-01 00:24:07	2019-12-01 00:44:40	2.0	3.2
103	1.0	2019-12-01 00:14:00	2019-12-01 00:18:46	2.0	0.9
107	1.0	2019-12-01 00:08:57	2019-12-01 00:20:32	2.0	3.1
235	1.0	2019-12-01 00:24:05	2019-12-01 00:36:19	2.0	1.0
335	1.0	2019-12-01 00:38:43	2019-12-01 00:44:23	2.0	1.1
...
6845088	1.0	2019-12-31 23:18:45	2019-12-31 23:27:02	2.0	1.9
6845217	2.0	2019-12-31 23:30:06	2019-12-31 23:43:57	2.0	1.3
6845225	1.0	2019-12-31 23:21:27	2019-12-31 23:33:01	2.0	1.9
6845231	2.0	2019-12-31 23:34:20	2019-12-31 23:58:23	2.0	2.2
6845267	2.0	2019-12-31 23:57:45	2020-01-01 00:30:13	2.0	17.5

81260 rows × 18 columns

3.2.b Total single riders being picked up in locationIDs from 0-50


```
In [84]: num_double_riders_0_50 = len(double_riders_0_50.index)
print("There were {0} total rides with double riders in being picked up at locationIDs from 0-50 in December 2016.".format(num_double_riders_0_50))
```

There were 81260 total rides with single riders in yellow taxis in December 2016.

3.2.c Total single riders being picked up in locationIDs from 0-50

```
In [85]: double_riders_0_50 = double_riders_0_50.loc[double_riders_0_50['DOLocationID'] > 0]
double_riders_0_50 = double_riders_0_50.loc[double_riders_0_50['DOLocationID'] < 50]
double_riders_0_50
```

Out[85]:

	VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance
463	1.0	2019-12-01 00:54:07	2019-12-01 01:18:01	2.0	12.1
836	2.0	2019-12-01 00:13:48	2019-12-01 00:17:36	2.0	0.7
1331	2.0	2019-12-01 00:34:07	2019-12-01 00:46:43	2.0	4.2
1558	2.0	2019-12-01 00:41:19	2019-12-01 00:45:39	2.0	1.0
2068	1.0	2019-12-01 00:37:41	2019-12-01 00:39:17	2.0	0.1
...
6843934	2.0	2019-12-31 23:55:58	2019-12-31 23:59:35	2.0	0.7
6844363	2.0	2019-12-31 23:34:24	2020-01-01 00:09:35	2.0	11.8
6844426	1.0	2019-12-31 23:30:56	2019-12-31 23:38:21	2.0	1.0
6844917	2.0	2019-12-31 23:06:53	2019-12-31 23:10:19	2.0	0.0
6844918	2.0	2019-12-31 23:06:53	2019-12-31 23:10:19	2.0	0.0

12789 rows × 6 columns

3.2.d Total single riders being picked up AND dropped off in locationIDs from 0-50

```
In [86]: num_double_riders_0_50 = len(double_riders_0_50.index)
print("There were {0} total rides with double riders in being picked u
p AND dropped off at locationIDs from 0-50 in December 2016.".format(n
um_double_riders_0_50))
```

There were 12789 total rides with double riders in being picked up A
ND dropped off at locationIDs from 0-50 in December 2016.