Drivers Analysis

Tal Mizrahi

Assumptions

- App measures the distance of a driver only when he's Online or Busy.
- If a driver update busy status while driving for other company, he's more engaged than another driver with the same online duration and online distance which doesn't update busy status.
- It's more accurate to analyze data without drivers that joined during the period (if full week is FALSE).
- I appreciate that Taxi drivers worked less than usual in the week between 22-29/12 because of the Christmas Eve holiday. Therefore, I didn't compare the daily distance of a taxi driver relatively to average in NYC in this period. I compared it on date 30/12/2015.

Analysis decisions

- I made two comparations:
  1. Includes all the driver's data that were active between 30-22/08/2015.
     Pros: The longest period in the database.
     cons: A lot of drivers aren't included for the reason they have started to drive after 22/08/2015
  2. Includes all the driver's data that were active in the last two days 29-30/12/2015.
     Pros: This period, has the most active drivers.
     cons: Data is for one day only and not for a longer period.
- To separate last-week columns. Last-week columns includes both dates: 29/12/2015 and 30/12/2019. Separation of 30/12/2015 essential to group correctly all the data together from 22/12/2019-30/12/2019.
- For most of charts, I have decided to include just 22-30/12 data for the reason there aren't many differences between them.

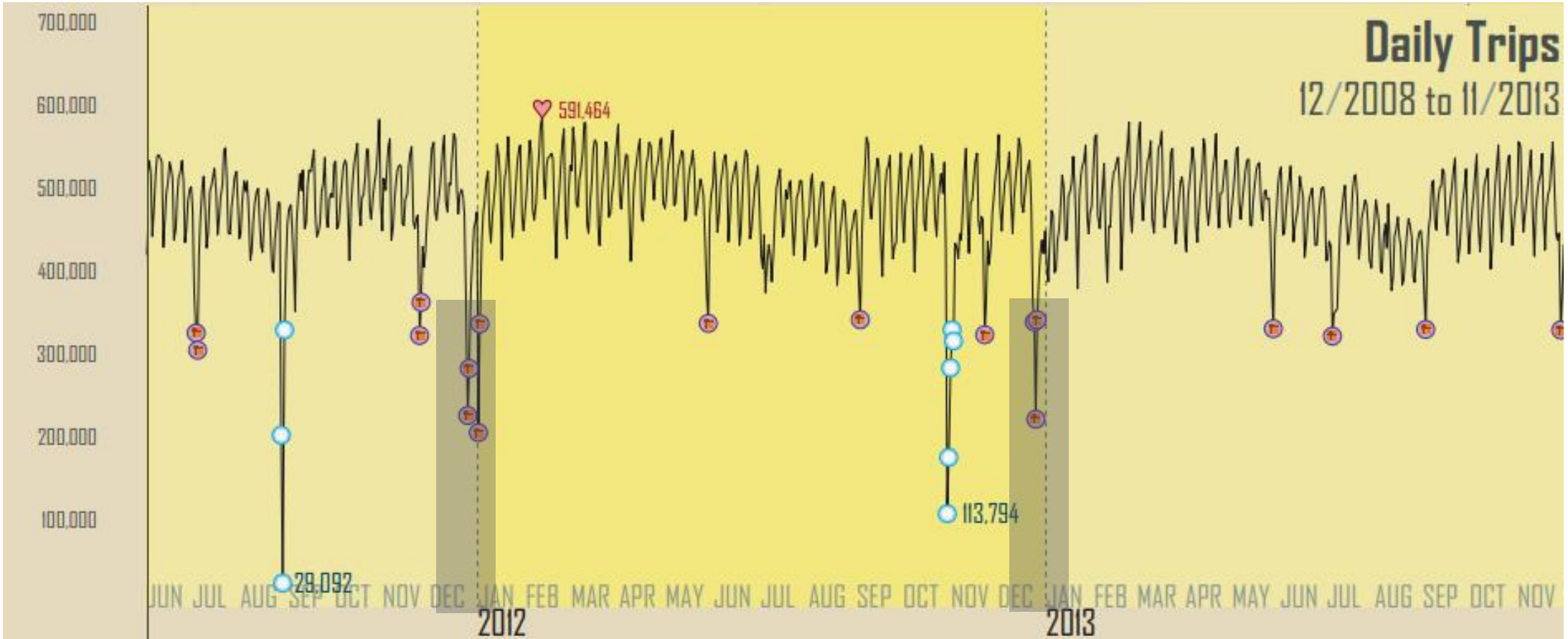
Filtered Data

The way I used to filter the data:

```
Counting number of drivers that are in both period times (29-30/12/2015 and 22-30/12/2015)

----------------------------Total Number of Drivers-------------------------------------
SELECT ProviderID FROM drivers GROUP BY ProviderID -- 260

--------------------------------- 22-30/12/2015-----------------------------------------
SELECT ProviderID
FROM drivers
WHERE ProviderID NOT IN (SELECT ProviderID FROM drivers WHERE Fullweek='FALSE')   //Filter full week
    AND BusyDuration + onlineDuration  != 0 //Filter N/A or non active drivers
    GROUP BY ProviderID --Total Drivers 146


--------------------------------- 29-30/12/2015-----------------------------------------
SELECT ProviderID
FROM drivers
WHERE recordtype = 'Day'
    AND ProviderID NOT IN (SELECT ProviderID FROM drivers WHERE recordtype ='Day' And Fullweek='FALSE')
    AND BusyDuration + onlineDuration  != 0
    GROUP BY ProviderID --Total Drivers 185
```

Time period

Time period of data is during Christmas Eve (dated in 24-25/12/2015) which may effected the data :



Engagement levels

I used two measures to conduct the engagement levels of a driver:

1. **Utilization** – measured by the ratio of busy time divided by total time (busy & online). As long as the driver spends more time on the online status, he's more engaged.
2. **Distance** – even if a driver spends less time on online status but spends a long time on busy status while making high distance relative to other drivers, it's valuable. Firstly he's a very good potential driver. Secondly, we can gather information and understand more about he needs and what he's looking for to understand what makes him use more other app or another service.

Ideal drivers for the company are which makes a lot distance with the highest utilization rate.

Engagement levels

There are 9 Engagement levels:

Utilization \ Distance	Low	Mid	High
Low	Low-Low	Low-Mid	Low-High
Mid	Mid-Low	Mid-Mid	Mid-High
High	High-Low	High-Mid	High-High

Low - Mid Distance, Low – Mid Utilization – New drivers/Low engagement.

Low Distance, Mid – High Utilization - Drivers that may take advantage of 50$ weekly payment for online and busy durations of more than 30 hours.

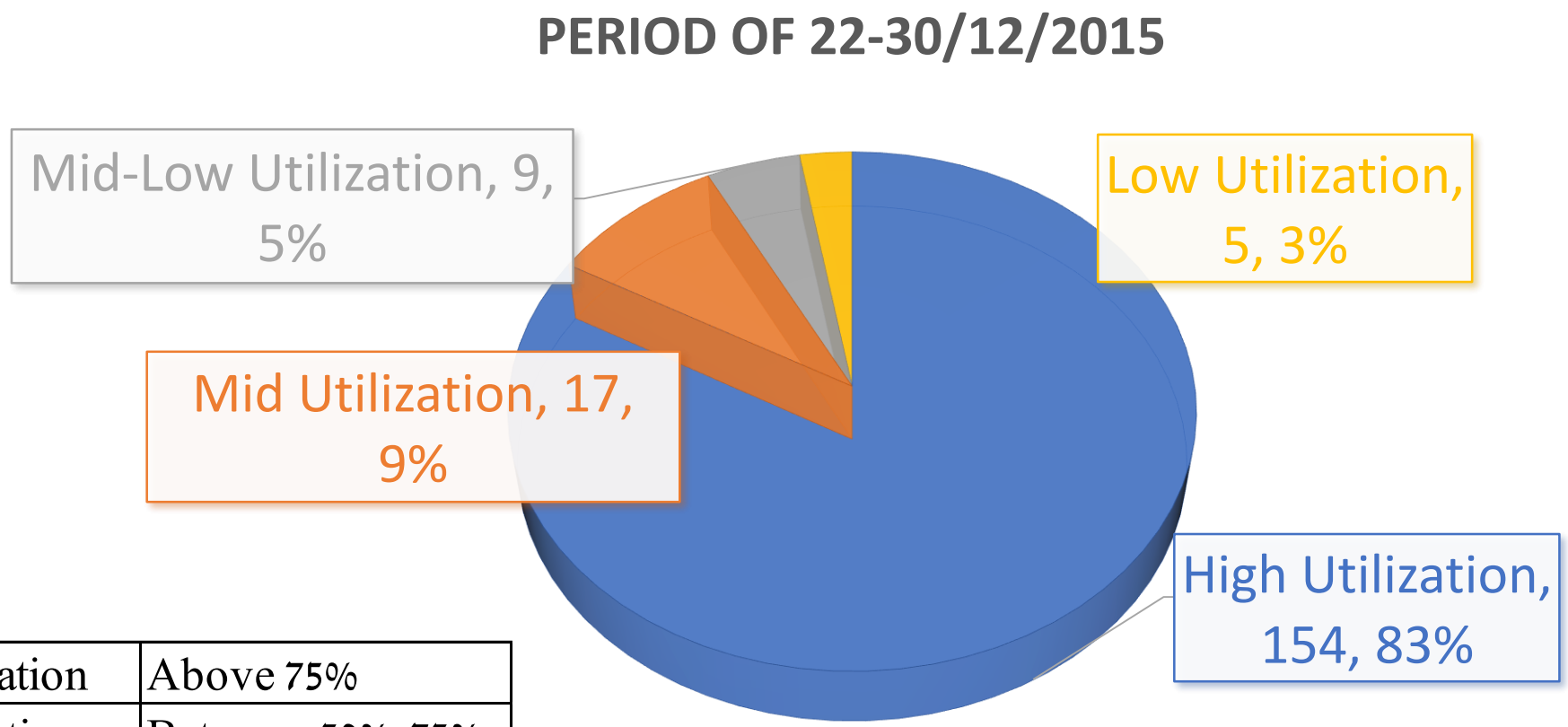
Mid – High Distance, Low Utilization – Drivers that have the highest potential of engagement.

High Distance, High Utilization – Best drivers.

Analysis description

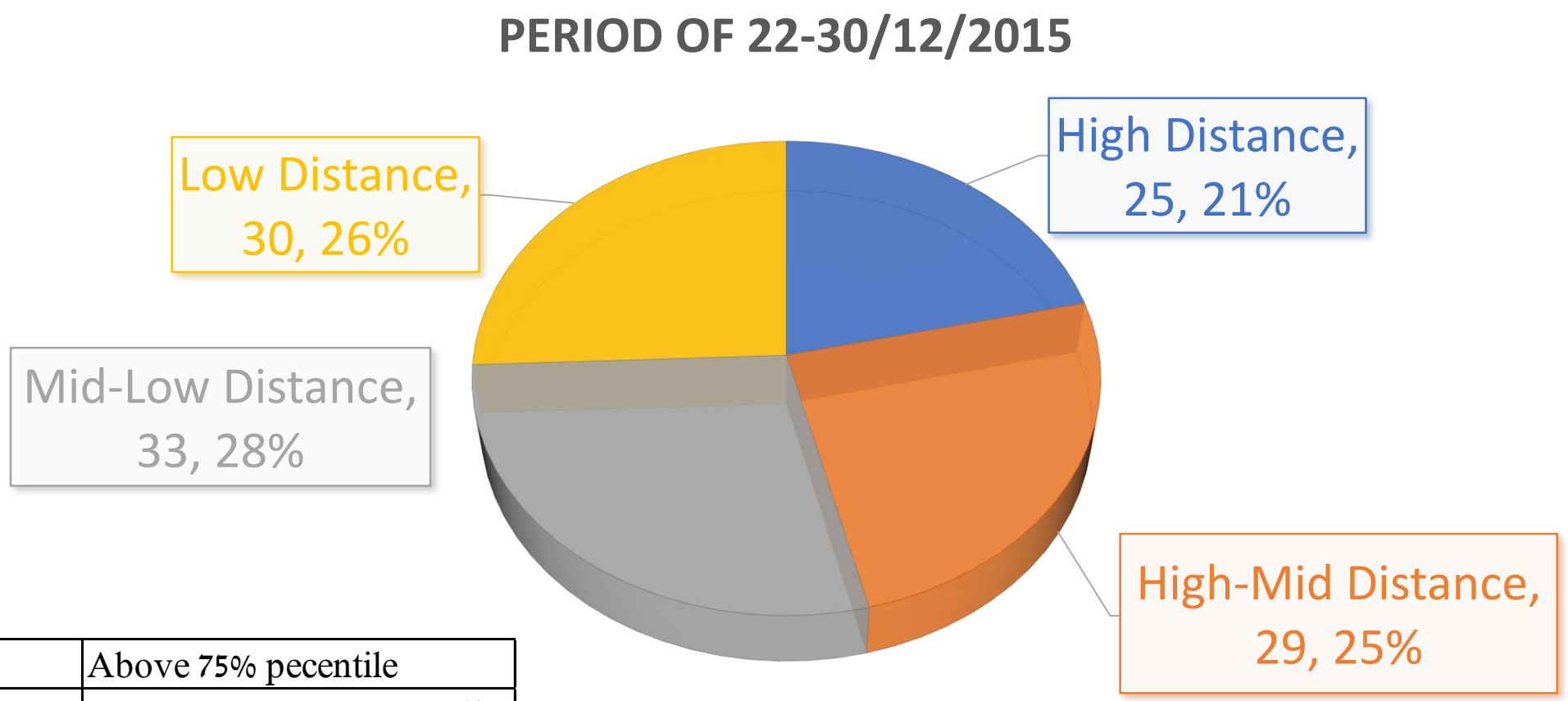
Chart Name	**Aim**
Utilization Distribution by Drivers	Measure engagement by Utilization
Distance Distribution by High Utilization drivers	Measure engagement by Distance
Utilization Distribution by Drivers	Identify drivers by engagement levels
Total Time and Total Distance	To see overall trends and indentify extreme cases
Comparison Between Busy and Online period	
Mean and Average of distance of drivers with Top total distance	
Beyond and below industry average by Utilization	Measure distance of drivers by indusry average daily distance

Utilization Distribution by Drivers



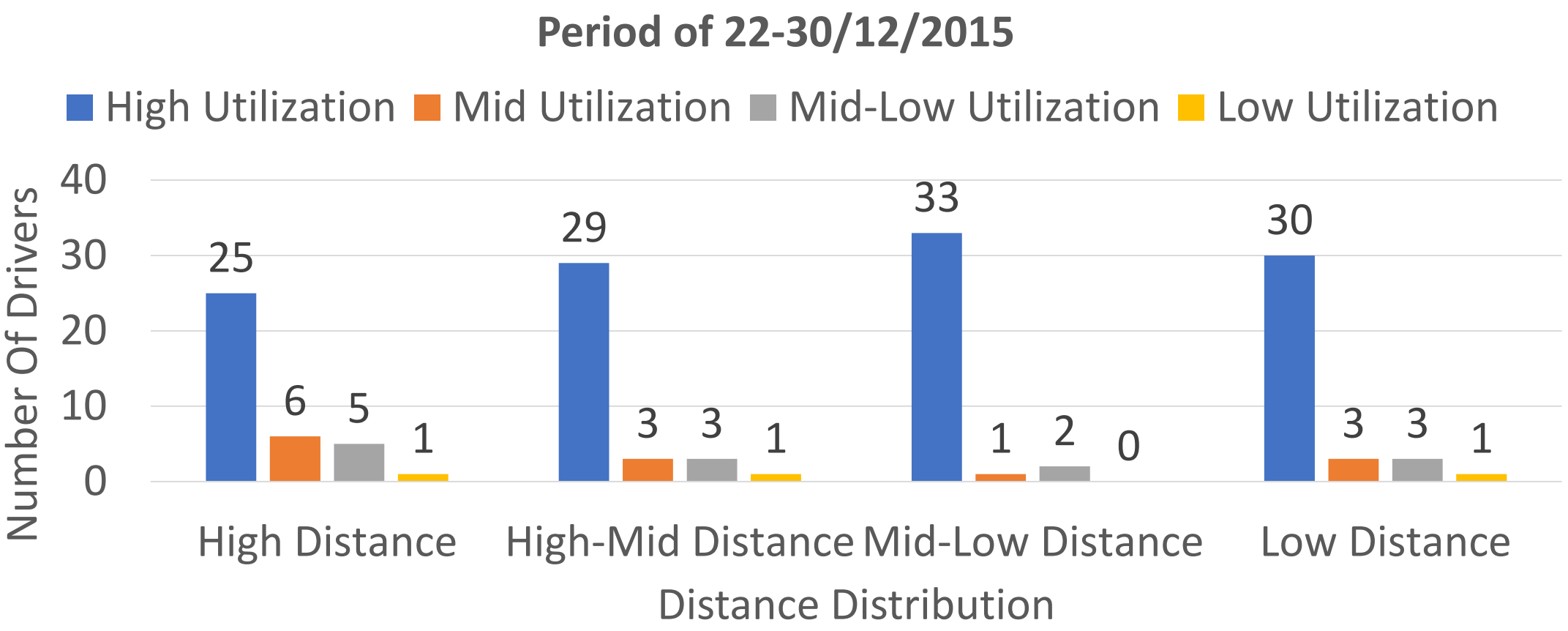
High Utilization	Above 75%
Mid Utilization	Between 50%-75%
Mid-Low Utilization	Between 25%-50%
Low Utilization	Below 25%

Distance Distribution by High Utilization drivers

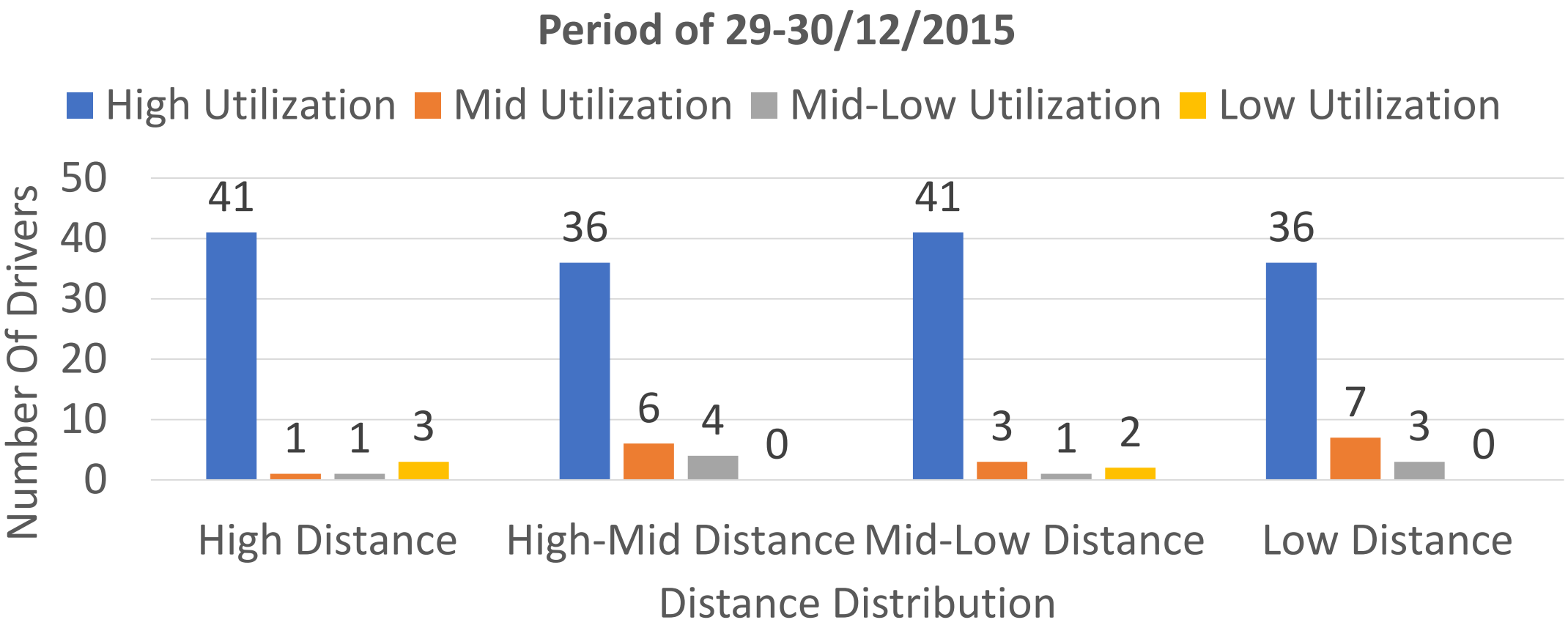


High Distance	Above 75% pecentile
High-Mid Distance	Between 50%-75% pecentile
Mid-Low Distance	Between 25%-50% pecentile
Low Distance	Below 25% pecentile

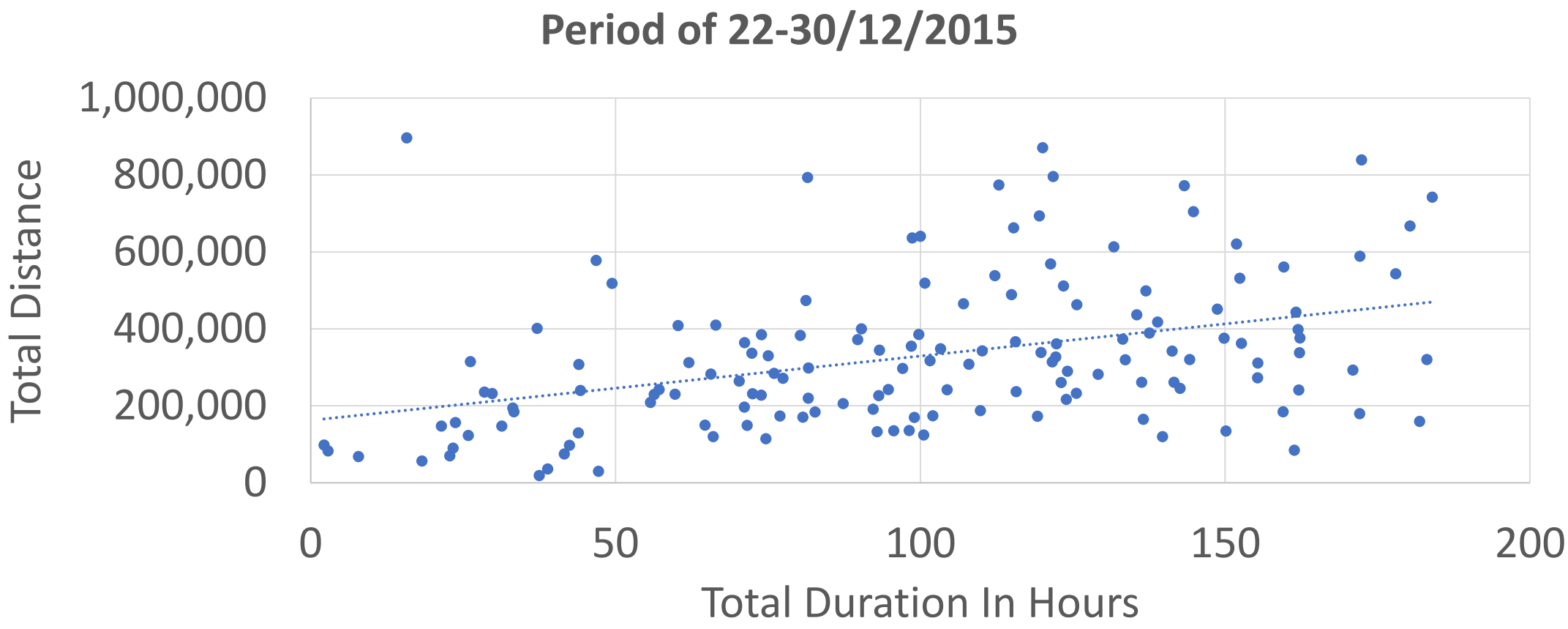
Utilization Distribution by Drivers



Utilization Distribution by Drivers



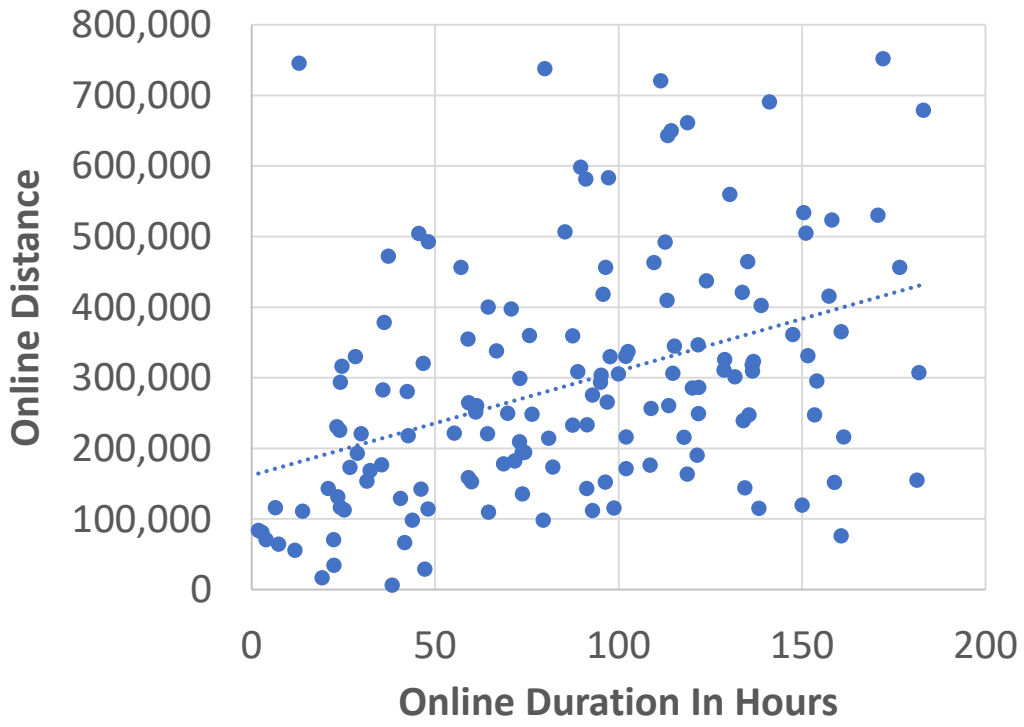
Total Time and Total Distance



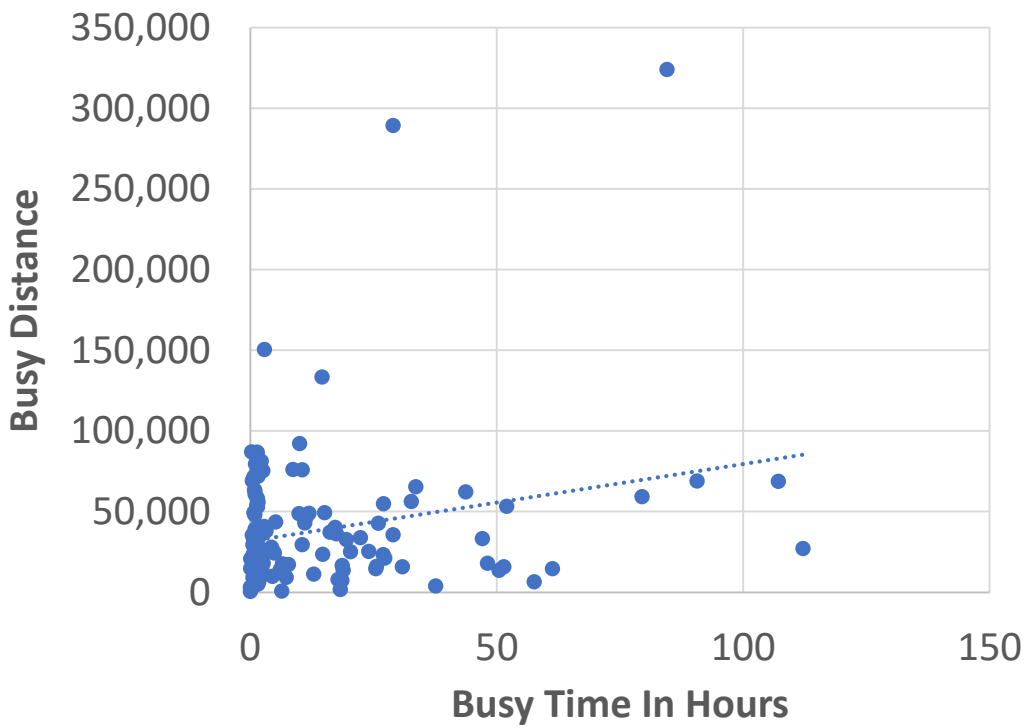
Comparison Between Busy and Online period

Period of 22-30/12/2015

Online Distance and Online Time



Busy Distance and Busy Time



Mean and Average of distance of drivers with Top total distance

ProviderID	Busy Distance	Online Distance	Total Distance	Distance Percentile	Avg Distance (Total Updates/Total Distance)	Mean Of Period 22-30/12	Mean Of Period 29/12
095d4caa-8e1f-4892-81ad-6309cdc8f6a8	150452.3	745792.1	896,244	1	28.38	31.3	0
85fa3519-ce82-4174-a836-f6c7b14ac185	289195.3	581533.6	870,729	0.993103448	119.67	8.5	6.4
59ea2233-87a6-4c7f-99ff-4fd5fd220f9e	86956.3	752027.1	838,983	0.986206897	46.69	6.6	7.7
3210d906-2e49-4eab-a3d7-99dea0a73b23	323947.2	472139.8	796,087	0.979310345	32.92	0.5	17651.8
156925ee-f81a-44da-b181-c29b0ee01262	55785.8	737830.1	793,616	0.972413793	22.60	2.4	7.5
7c5402e4-0024-4a39-8215-b557aed3a010	53301.4	720758.4	774,060	0.965517241	10.76	2.9	6.8
4fb70d24-4795-4c3c-a1bf-2ea3bf87a310	81310.9	690815.1	772,126	0.95862069	16.09	1.1	1.4
669f37a5-e52f-4e00-9db0-b70a01ea60c8	62778	679240.3	742,018	0.951724138	24.25	11.1	1.4
4d21e39c-8e83-4905-aeb4-1f5c00d5ca77	42831	661540.6	704,372	0.944827586	51.94	1.1	0.5
7c93b300-031a-4ed0-a8f5-fc48b7843969	43543.7	649838.6	693,382	0.937931034	18.16	0.8	10.6

NYC taxi drivers average distance per day

Accordance to the FACT book of 2014[2] : "A typical taxi travels 70,000 miles per year".

I have made a calculation to compare the data of taxi drivers to the industry.

It may not be accurate because of 2 reasons:

1. Best to measure the exact period.
2. My assumptions of the calculation.
3. A Short period measured

Calculation method

70,000 are 112,685 Kilometers.

112,685/365 * 1,000 / 2 = 154,363 meters

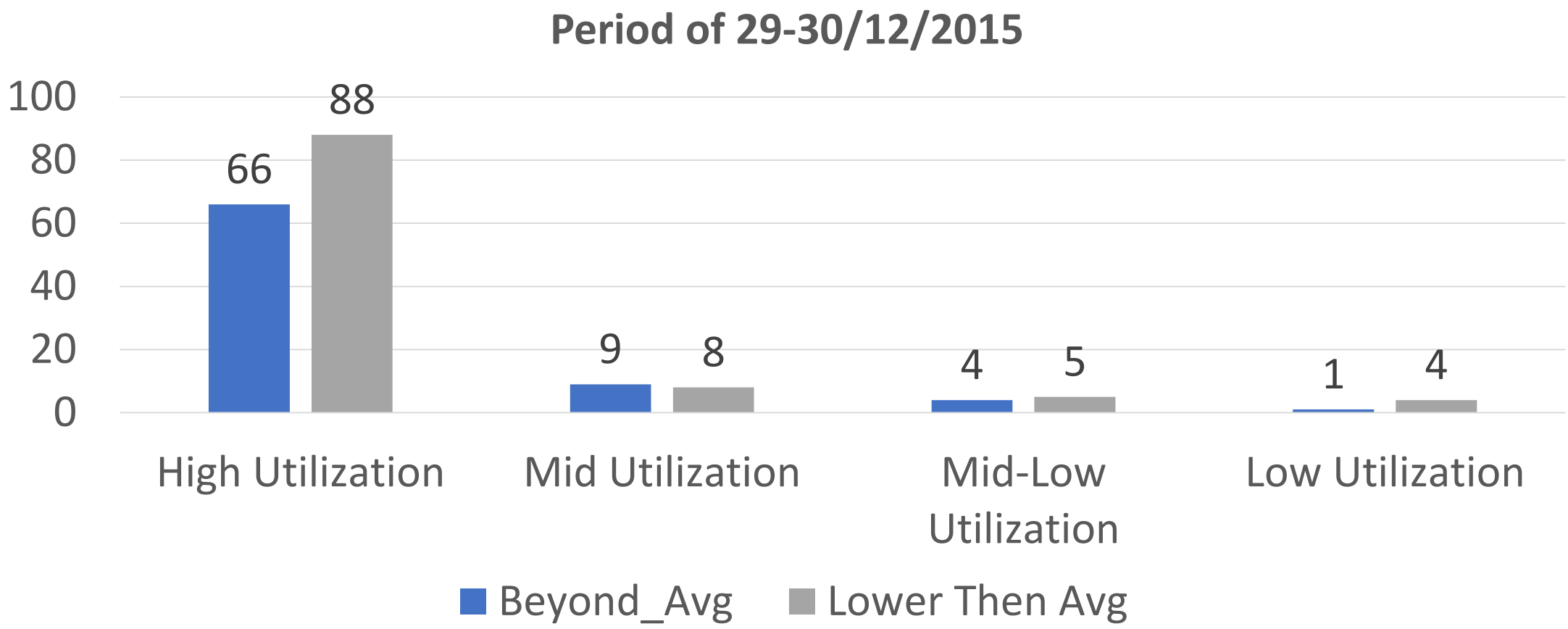
365 – Days in a year

2 – Number of shifts per day in a taxi

1,000 – conversation from kilometers to meters

2. source: **Factbook TLC NYC 2014**

Beyond and below industry average by Utilization



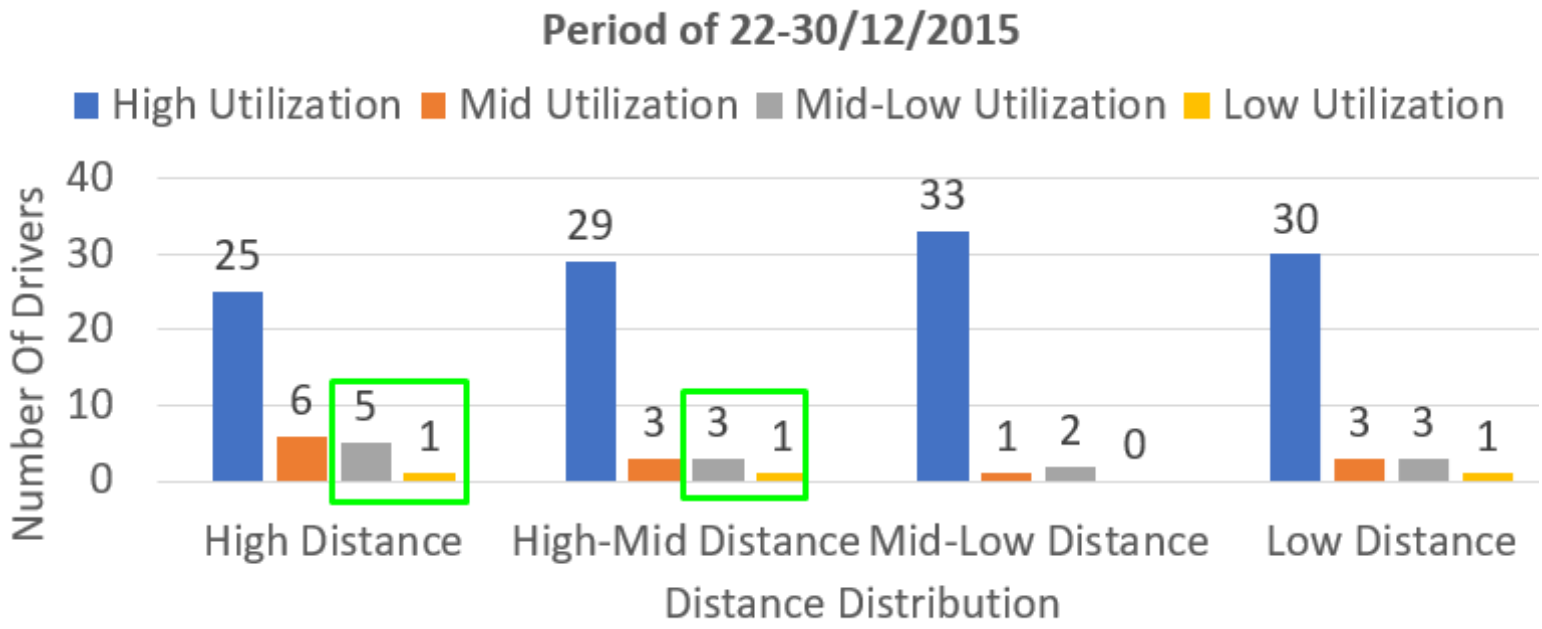
Conclusions

- Over 80% of the drivers are in an online duration of at least 75% of total time (online + busy duration). It indicates a very high engagement of drivers.
- Drivers with high utilization aren’t necessarily engaged to drive more distance (A conclusion from Utilization Distribution by Drivers charts and comparison between them and the industry average)
- There are drivers which take advance of 50$ weekly payment. They are on active and busy duration more than 30 weekly hours, however they drive very little distance.
- Drivers with the highest distance among the other drivers, used to do a lot of trips and not small numbers of trips with a long-distance.

Drivers that needs special attention

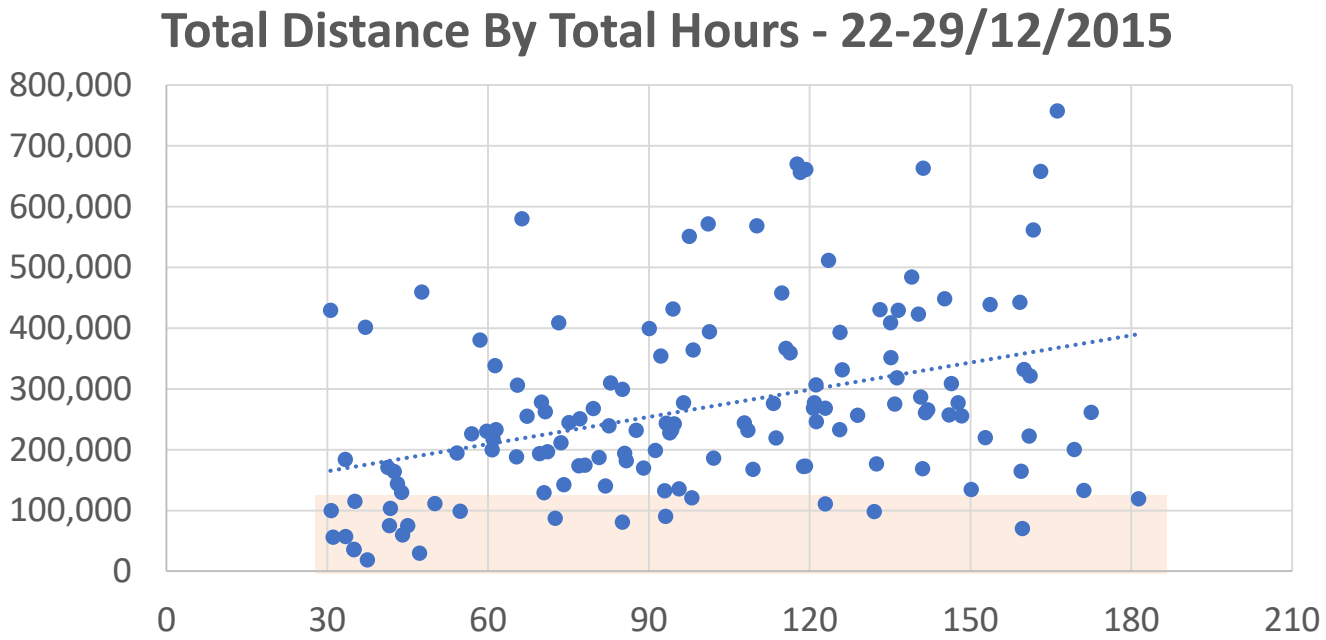
In my opining, the following drivers need special attention:

1. Drivers with low utilization and high distance. By conduct their behavior, we may understand why some drivers prefer other platforms.



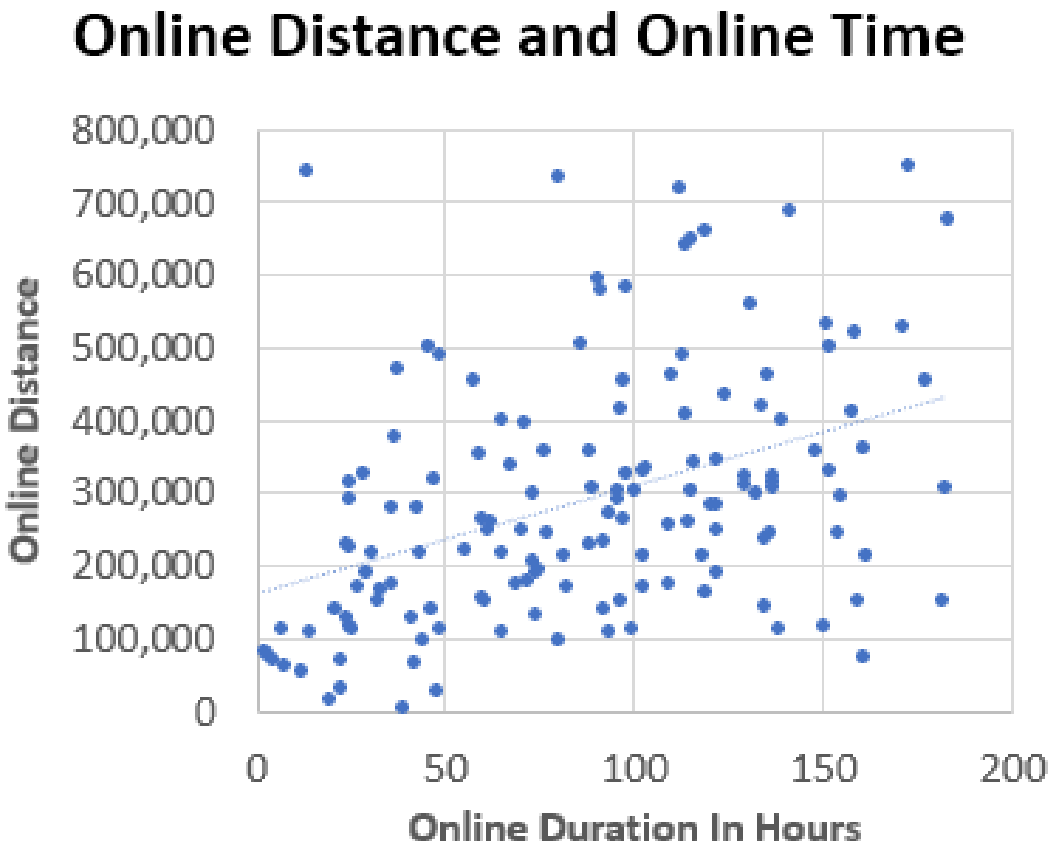
Drivers that needs special attention

2. Drivers with low distance and Total duration of online and busy higher than 30. Drivers may take advantage of it without taking a trip. I Have created a chart of 7 days (22-29/12/2015) to identify potential drivers:



Recommendations

- There is a high distribution between the drivers. I would recommend learning the differences between them
  to understand why
  some of the drivers drive much more then others and try to
  encourage them to drive more.



Recommendations

- In order to increase the distance, encourage drivers to drive in places with high demand, mostly short trips and in hours where there is less traffic.
- Set a distance requirement as well to receive 50$ weekly bonus.
- Ask for feedback from drivers with high distance and low utilization. It may be in a personal note with offers a driver that completes the feedback 3 next trips with a lower commission fee for example.