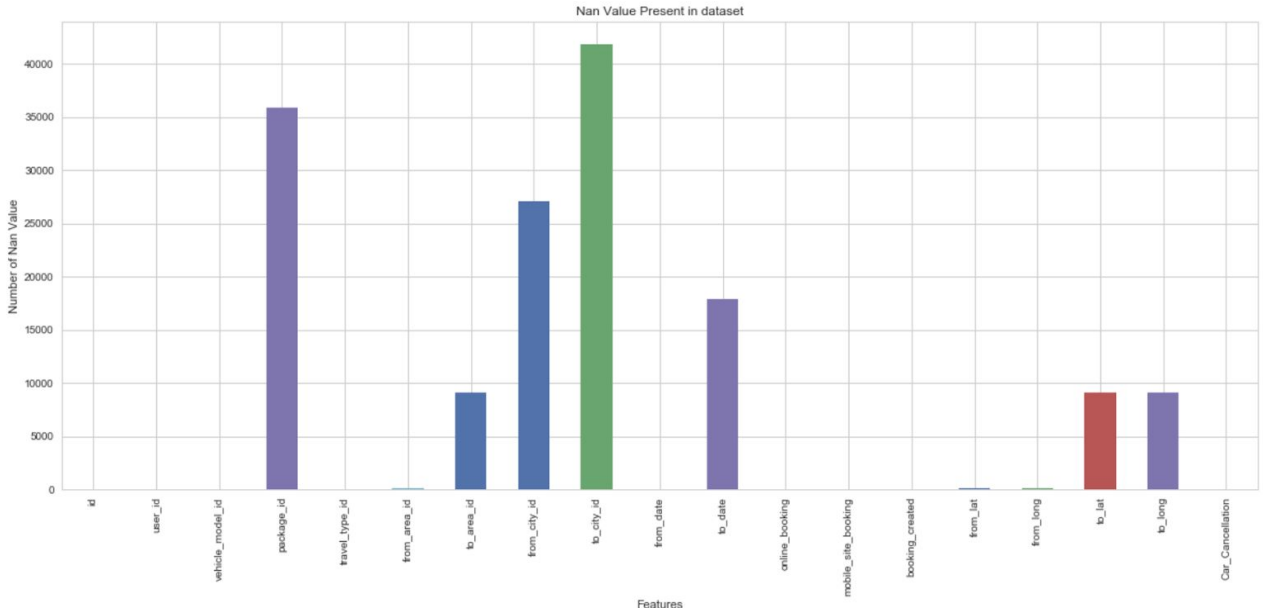
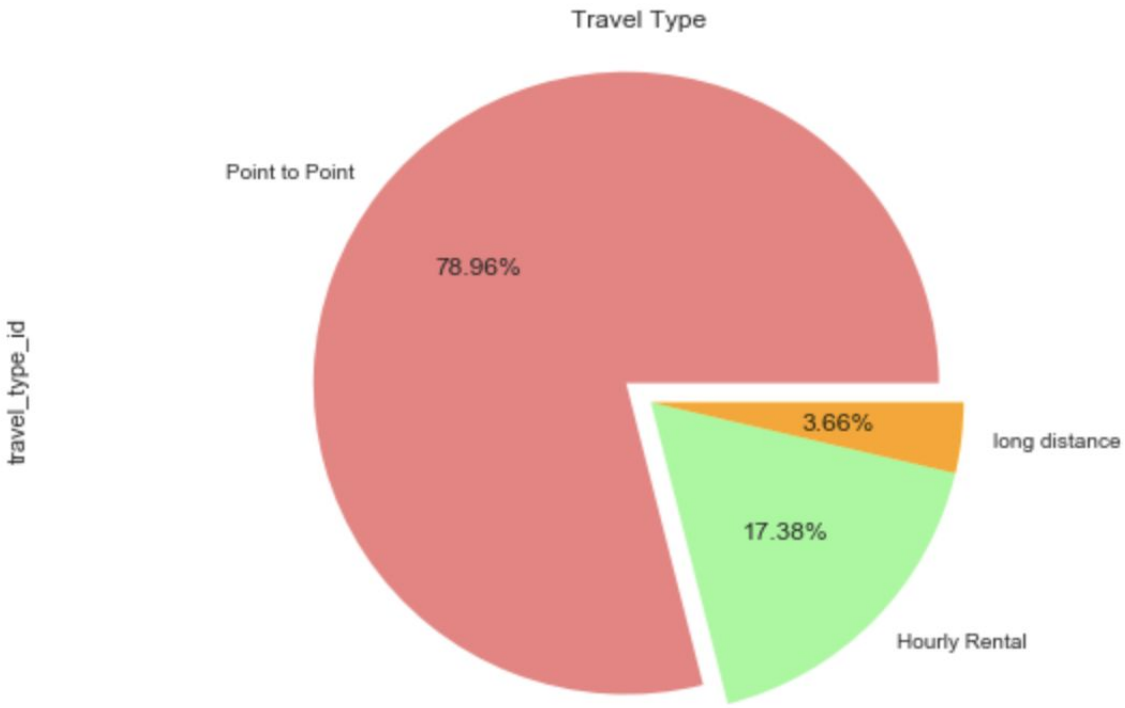
Analysis of Data Set

Check for the null values present in Dataset:

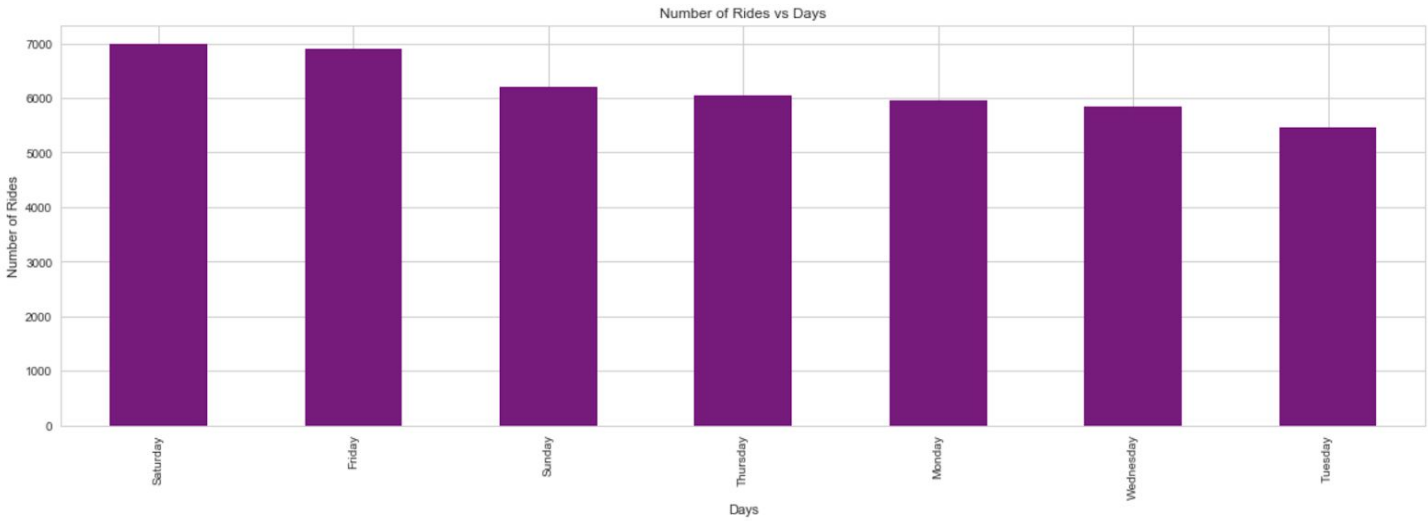


Feature	Percent of Null Value
package_id	82.6 %
from_area_id	0.20 %
to_area_id	21.04 %
from_city_id	62.36 %
to_city_id	96.34 %
to_date	41.19 %
from_lat	0.21 %
from_long	0.21 %
to_lat	21.04 %
to_long	21.04 %

The dataset is highly skewed towards the **Point to Point** travel



Here we find the number of rides taken vs Days (to analysis whether there is some trends or not)



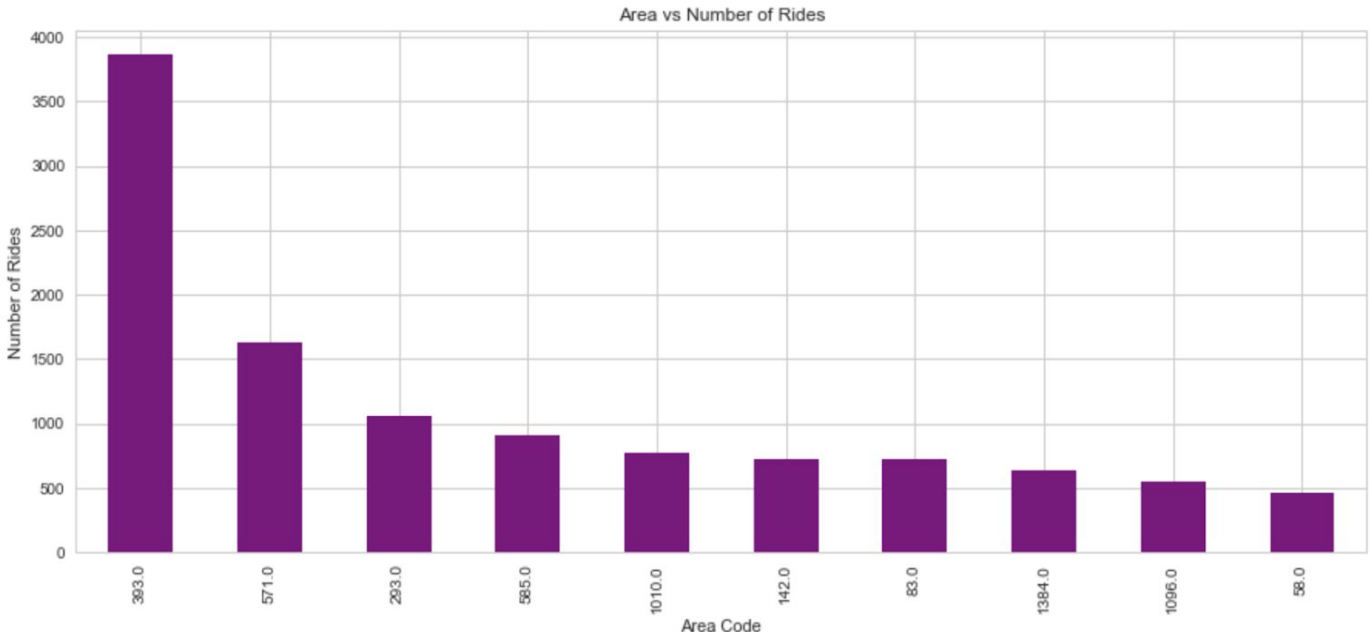
There is no specific trends in taking cab on specific day.

Lets analysis hour of the day and to find some unique trend whether there is some specific time when cabs demand goes high.
Before moving to analysis lets diving the area into four category

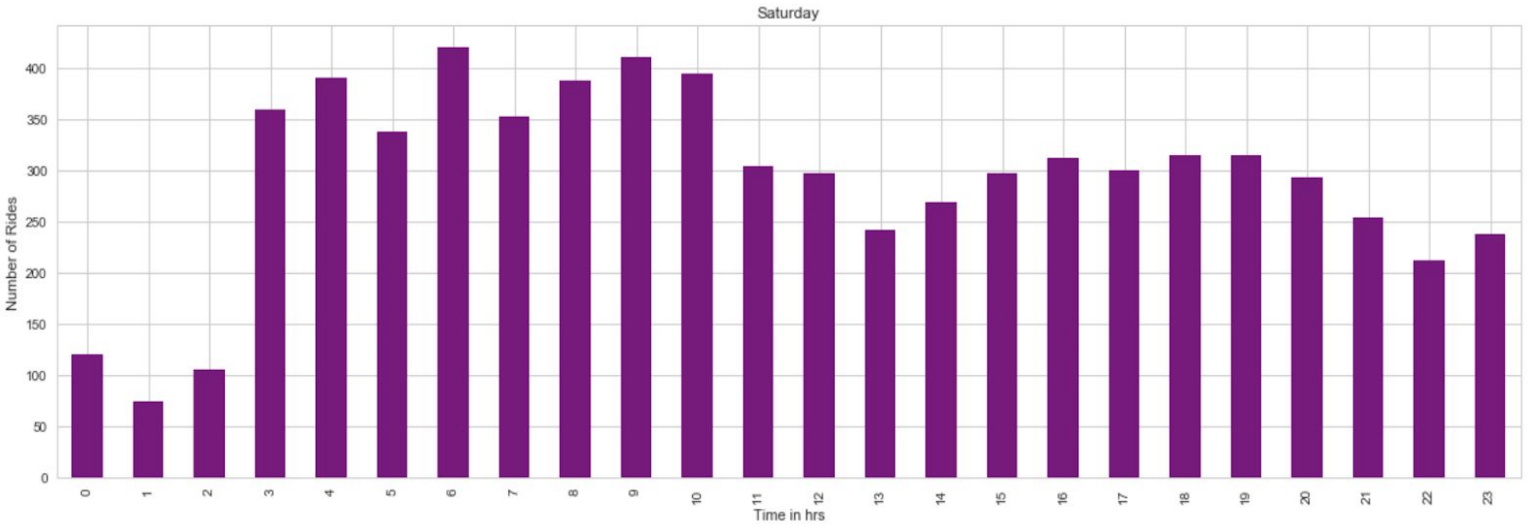
- **Public Transport**
- **Office Area**
- **Residential Area**
- **Market Area**

Dividing the category will help us to analyze the cab demand in a different area.
By the help of google search and feature from_lat and from_long find out the details of the area.
Public Transport:Airport(393),Bus Station(585)
Office Area: Marthahali(571),Electronic City(293),Brookefield(1096), Whitefield(1010),Arekere'(58)
Market Area : Yeswanthpur (1384)
Residential Area: BTM 2nd Stage(142), Bellandur(83), Yelahanka(1017), Dooravani Nagar(1371)

Only this area was considered since most of cabs demand come from this area.



Saturday:



From this graph, we find out that most of the cab demand in early morning and evening.

Lets us analysis the area wise cab demand on Saturday:

Graphs are already plotted on Jupyter Notebook File.

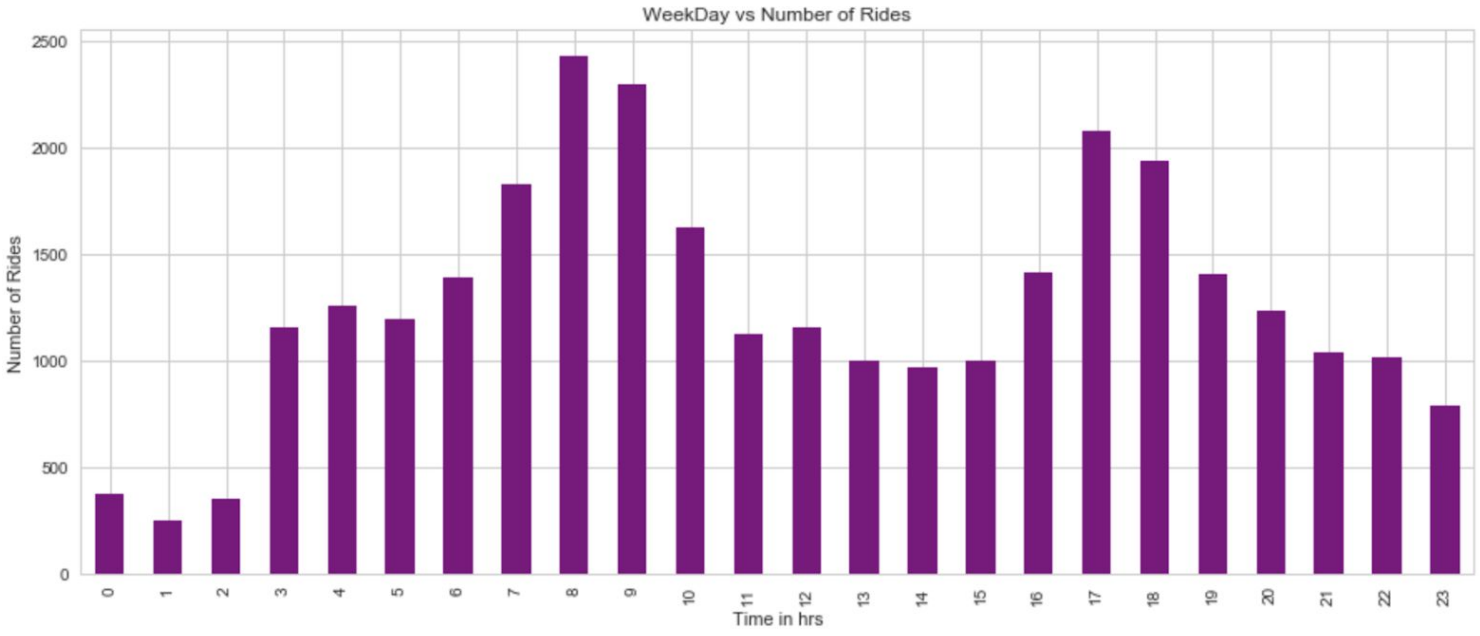
→ **For Office Area :**
Marthahali 3:00 AM -7:00 AM
Electronic City 6:00 AM - 8:00 AM
Brookefield 4:00 AM- 6:00 AM
Whitefield 9:00 AM-10:00 AM
The office area has a huge demand for Cabs in the morning.
→ **For Residential Area:**
Bellandur 3:00 -5:00
BTM 2nd Stage 3:00 - 5:00
Yelahanka 17:00 - 19:00
The residential area has very much uneven time for demands of cabs.
→ **For Market Area:**
Yeswanthpur: 20:00 - 22:00
The market area has a huge demand of cabs at 8:00 PM to 10:00 PM.

→ **For Transport Area**:
Airport: 8:00A.M - 10:00 A.M
Bus Station 6:00A.M - 8:00 A.M
The Transport area has a huge demand for Cabs in the morning.

Increase the supply of Cab in this specific area in this time-slot due to the high demand for cabs.

After analysis the data we find out the similar pattern of Cab demand on Weekdays i.e Monday - Friday.

Monday-Friday



From this graph, we find out that most of the cab demand in early morning and evening.

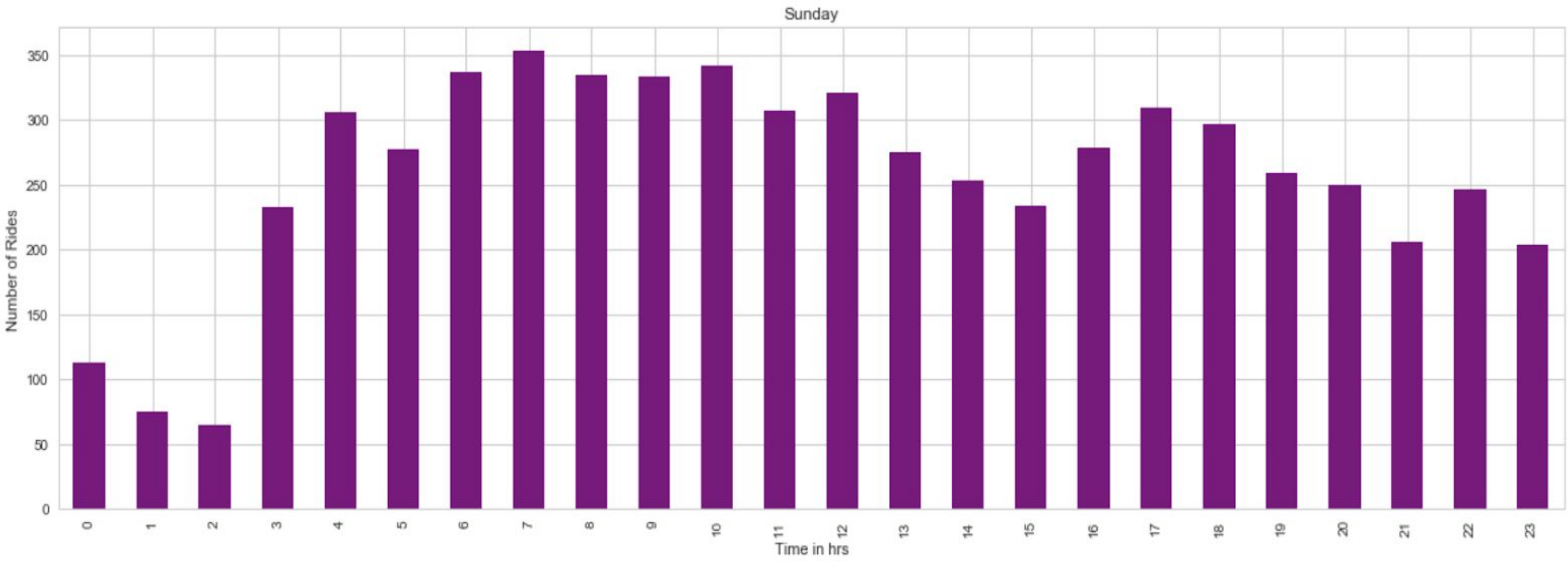
Lets us analysis the area wise cab demand on WeekDays:

Graphs are already plotted on Jupyter Notebook File.

→ **For Office Area :**

Marthahali 2:00 AM-3:00 AM and 8:00 AM-10:00 AM and 17:00 - 20:00
Electronic City 17:00 - 19:00
Brookefield 17:00 - 21:00
Whitefield 17:00 - 19:00
Arekere 7:00 AM - 10: 00 AM
The office area has a huge demand for Cabs in the morning and in the evening.
→ **For Residential Area :**
Bellandur 7:00 -9:00 and 18:00- 20:00
BTM 2nd Stage 3:00 - 5:00
→ **For Market Area:**
Yeswanthpur: 7:00 - 8:00
→ **For Transport Area:**
Airport: 22:00 - 0:00
Bus Station 22:00 - 23:00
The Transport Area has a huge demand of cabs at 10:00 PM to 00:00 AM.

Sunday:



On Sunday there is a holiday so there is so much demand for cabs, so there is no specific time.

Lets us analysis the area wise cab demand on Sunday:

Graphs are already plotted on Jupyter Notebook File.

→ **Office Area** :
Marthahali 6:00 - 8: 00
Electronic City 11:00 -12:00 and 16:00 - 19:00
Brookefield 17:00 - 21:00
Whitefield 16:00 -18:00
→**Residential Area** :
Bellandur 12:00 - 14:00
Dooravani Nagar: 4:00 - 5:00 and 6:00 - 7: 00
BTM 2nd Stage : Always Busy
→**Market Area** :
Yeswanthpur: 7:00 - 8:00
→ **Transport Area** :
Airport: 22:00 - 0:00
Bus Station 6:00 - 8:00

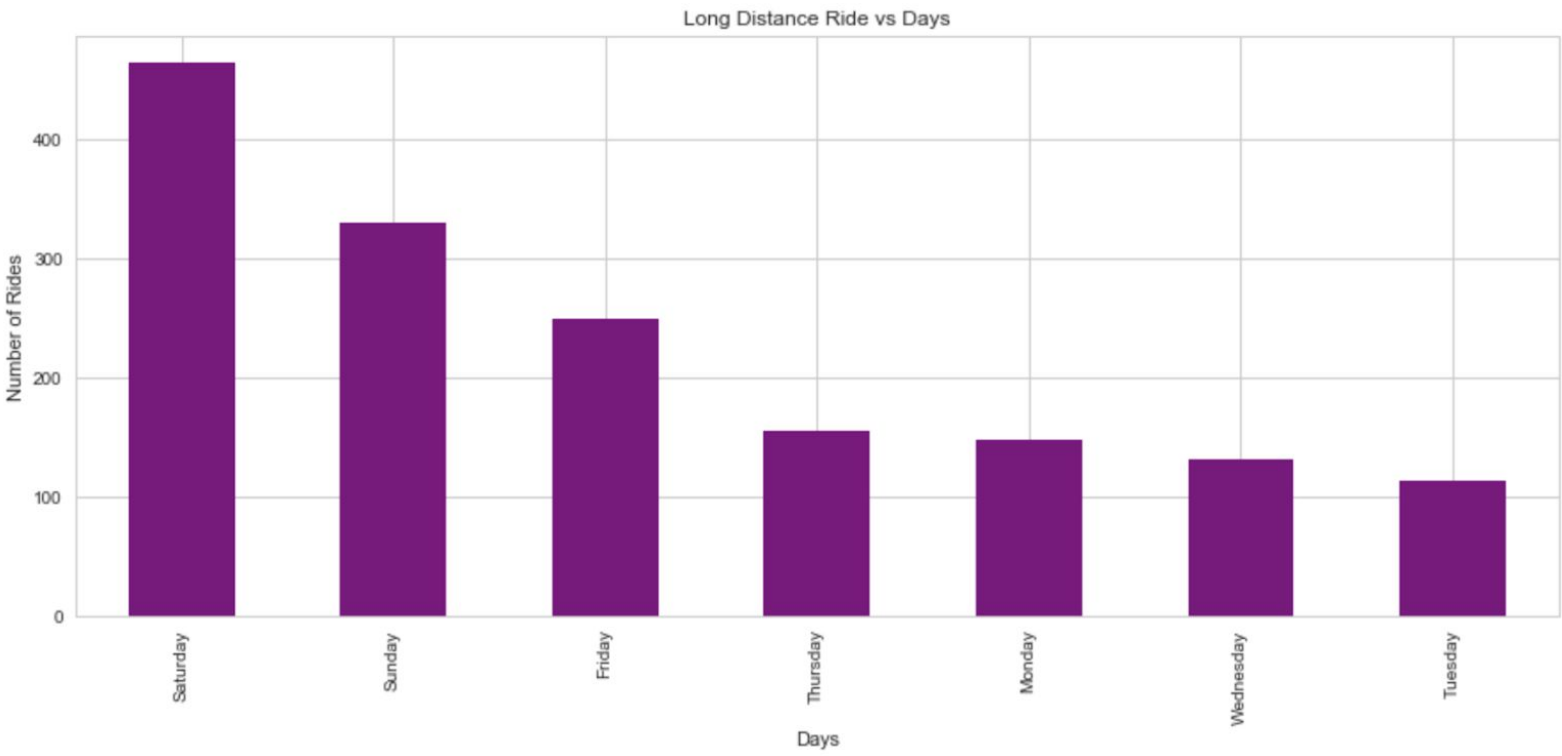
From the above analysis we found the important key Point:

⟶ The Weekday i.e Monday to Friday follows the same pattern.
→ In Weekday , the early morning has more traffic in the residential area and in the evening there is more traffic in the office area.
→ Saturday has more traffic in Market Area at 20:00 - 22: 00 .
→ Whole Data is divided into three categories:WeekDay, Sunday and Saturday. There is a specific time of traffic in the Transport Area in these three categories.
→ Here we develop a surge strategy and give incentives to the driver by doing surge pricing of cabs at peak hours.

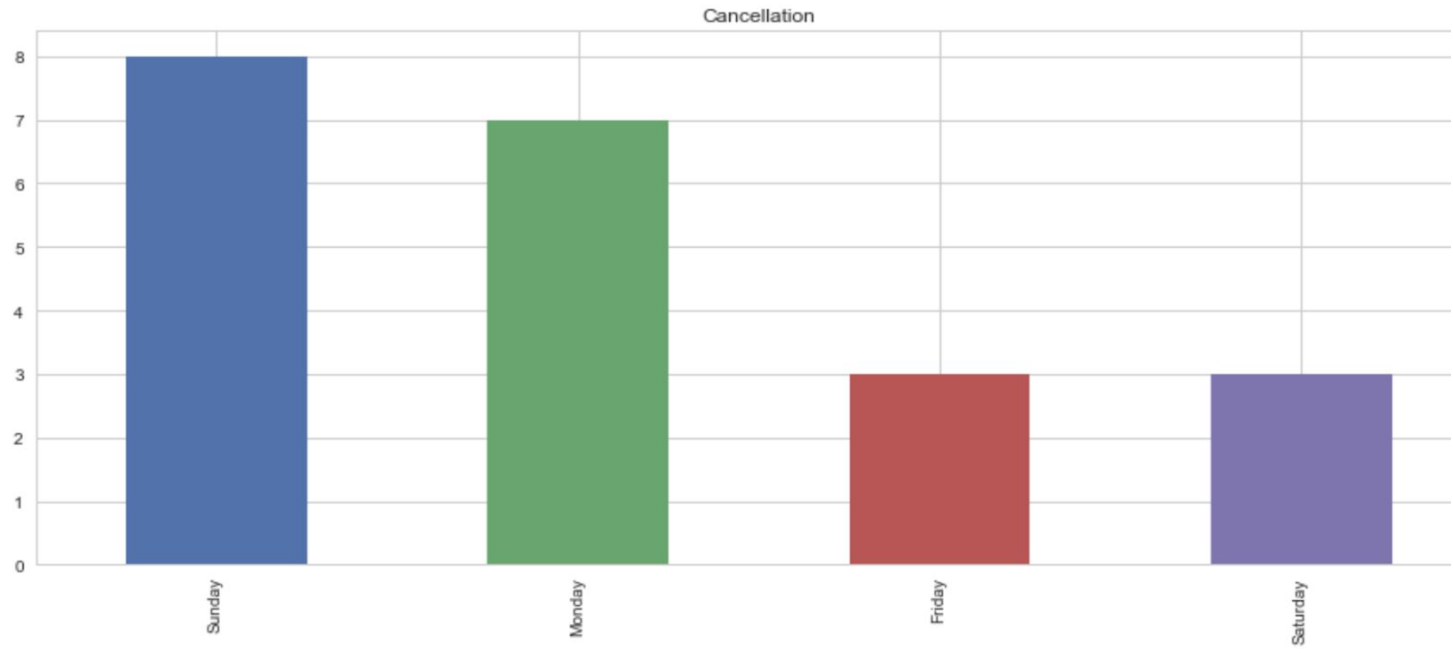
Travel Type:

- Long Distance Travel
- Point to Point Travel
- Hourly Rent Travel

Long Distance Travel:

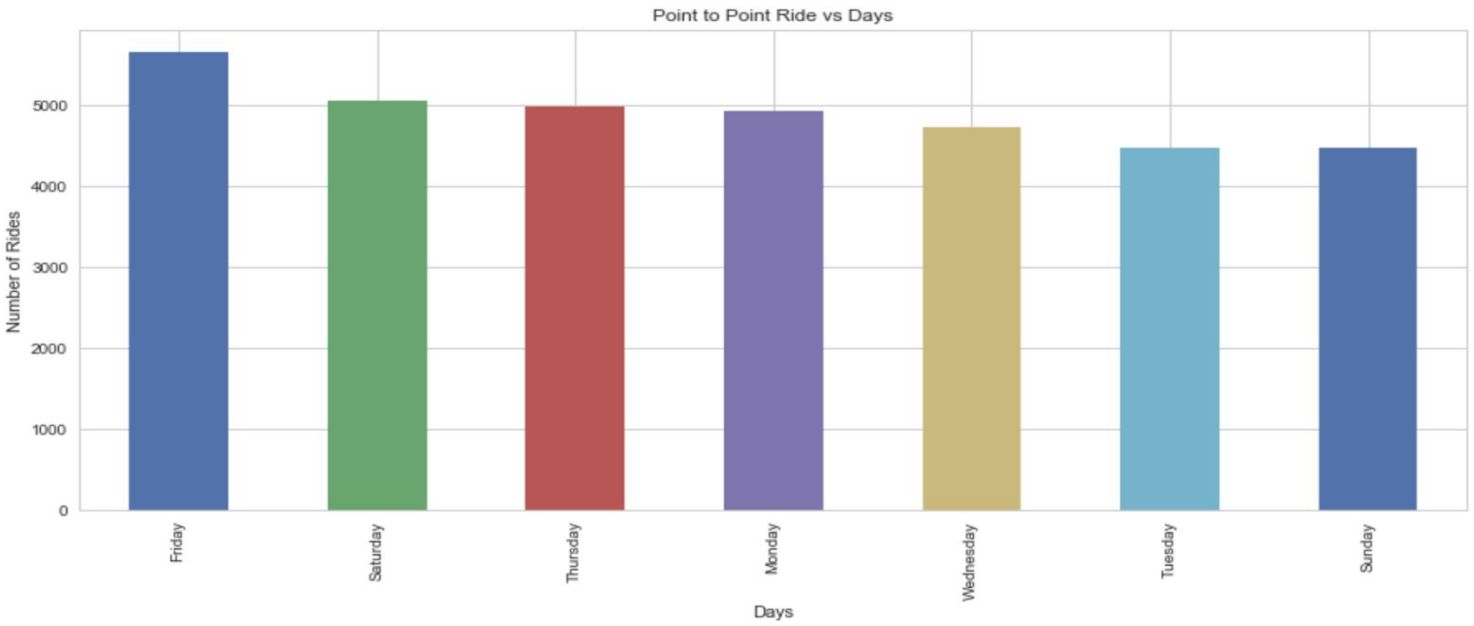


Cancellation:

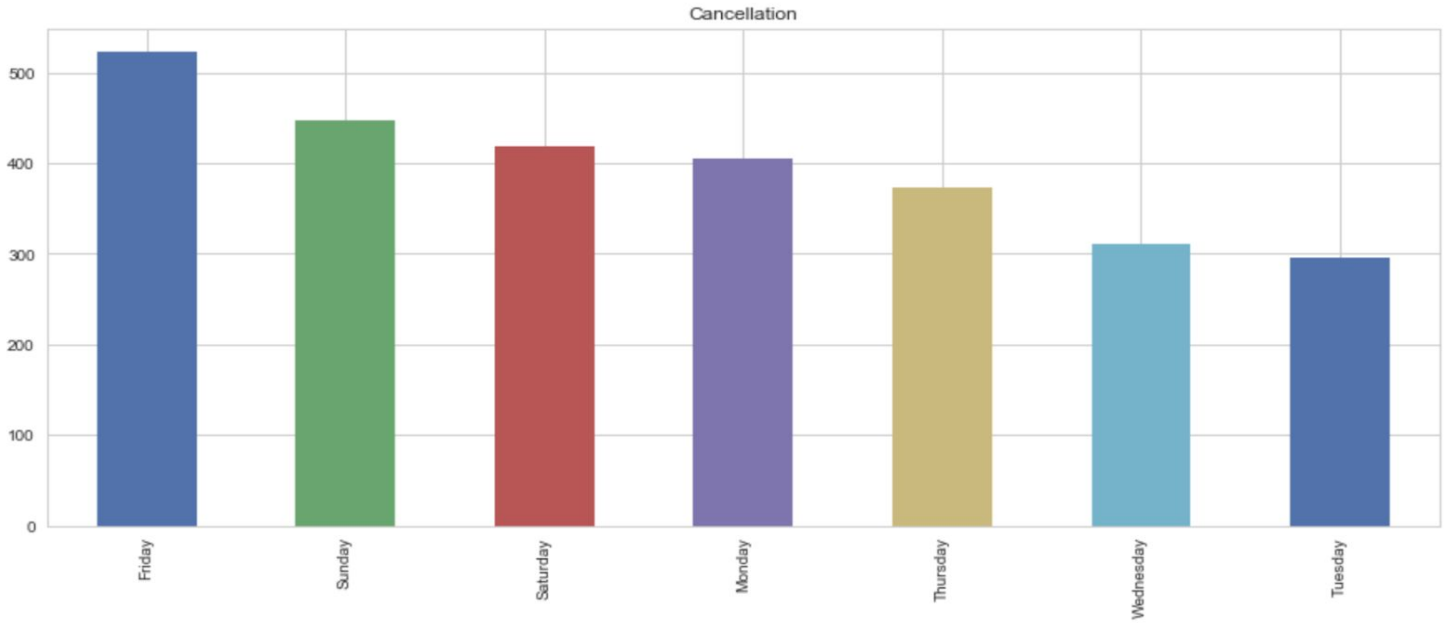


- Long Distance Travel is mostly done on weekends or on Friday Night
- Most of the Cancellation of Ride is done on Monday and Sunday.
- So the pricing of the cabs for the long distance should go up on weekends to meet the demand of customers and this will boost the incentives of driver

Point to Point Travel:

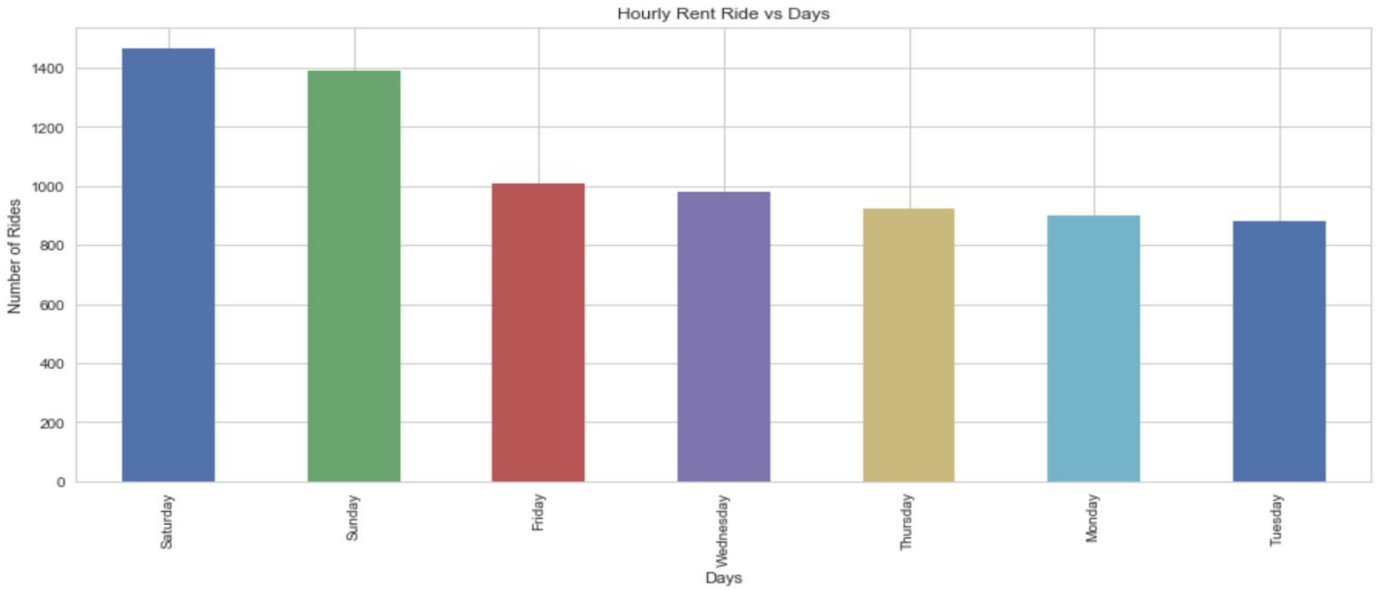


Cancellation:

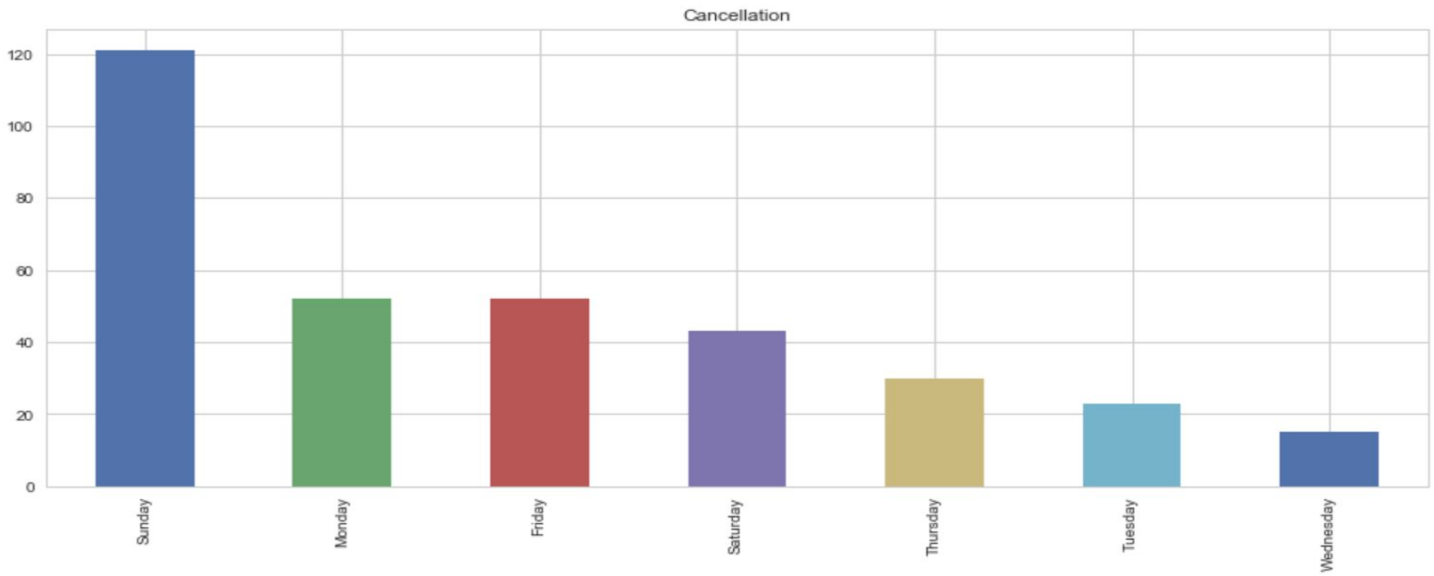


- Point to Point travel is generally used by office people on weekdays and Saturday.
- Here Point to Point Cancellation increases drastically on Sunday due to the shortage of cabs on Sunday. Since long distance and hourly rent takes the major cab share.

Hourly Rent Travel:

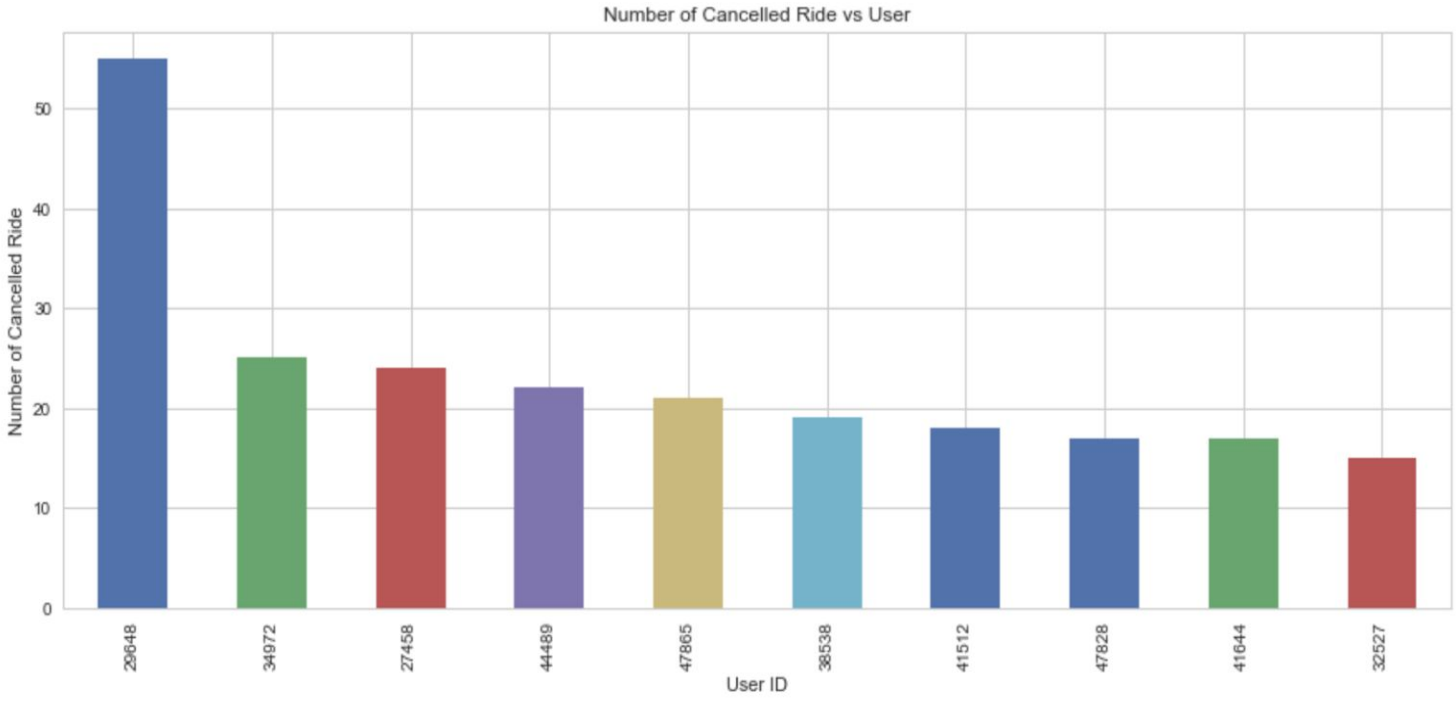


Cancellation:

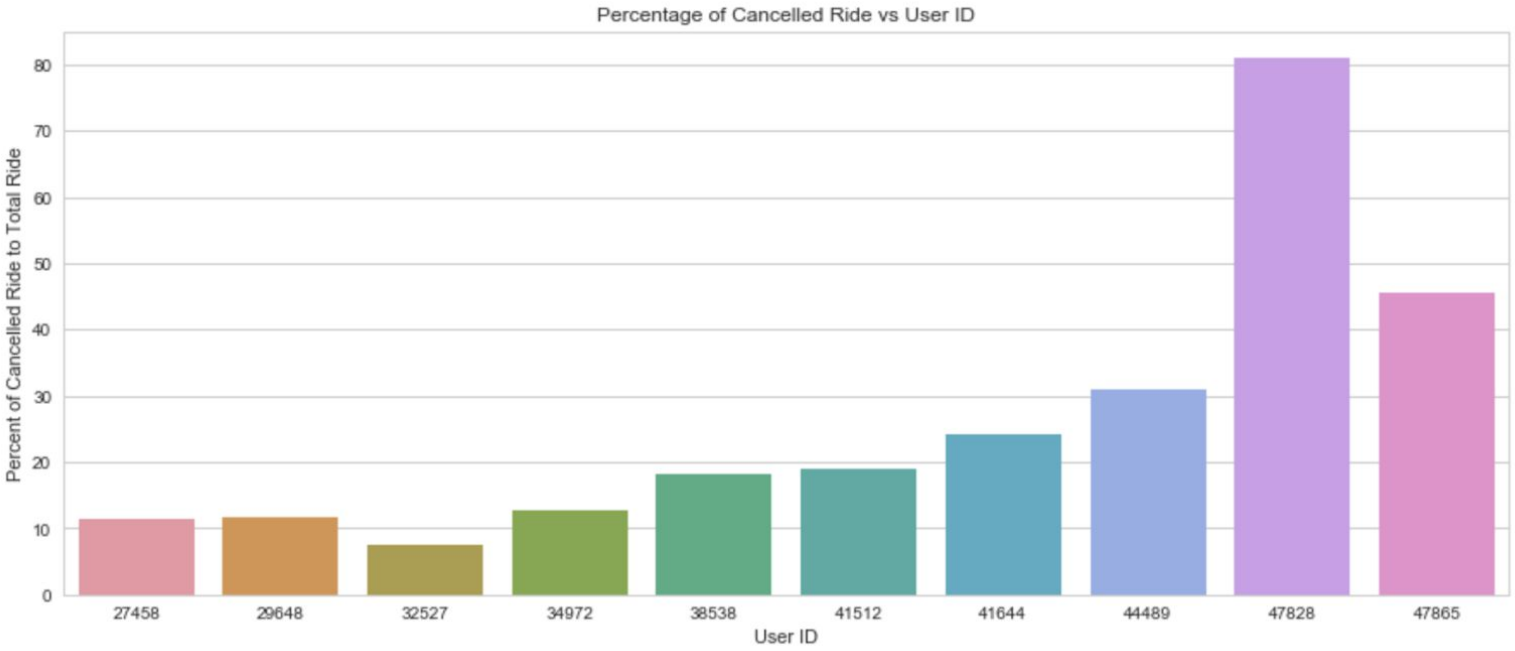


- Hourly Rent travel is mostly used during weekend
- Here Hourly Rent Cancellation increases drastically on Sunday due to the shortage of cabs on Sunday.

User Responsible for Car Cancellation



We can see that some user has the most number of the canceled ride.



Two User has more cancellation percent
User ID- 47828 : 81%
User ID- 47865: 46%
There are some doubtful user so cab company should take some action against them and run the analysis of whether user involved in some fraud or not. This will help us to increase the revenue.

For More Details Check the Jupyter Notebook file.