**Statistical Analysis on Bank Marketing Dataset**

**Name - Ritik R Mohapatra Student ID - 23065335GitHub -** [**https://github.com/Foxtrot123-png/Applied-Data-Science-1/blame/main/User\_Behaviour.ipynb**](https://github.com/Foxtrot123-png/Applied-Data-Science-1/blame/main/User_Behaviour.ipynb) (User Behavior Dataset, 2024)

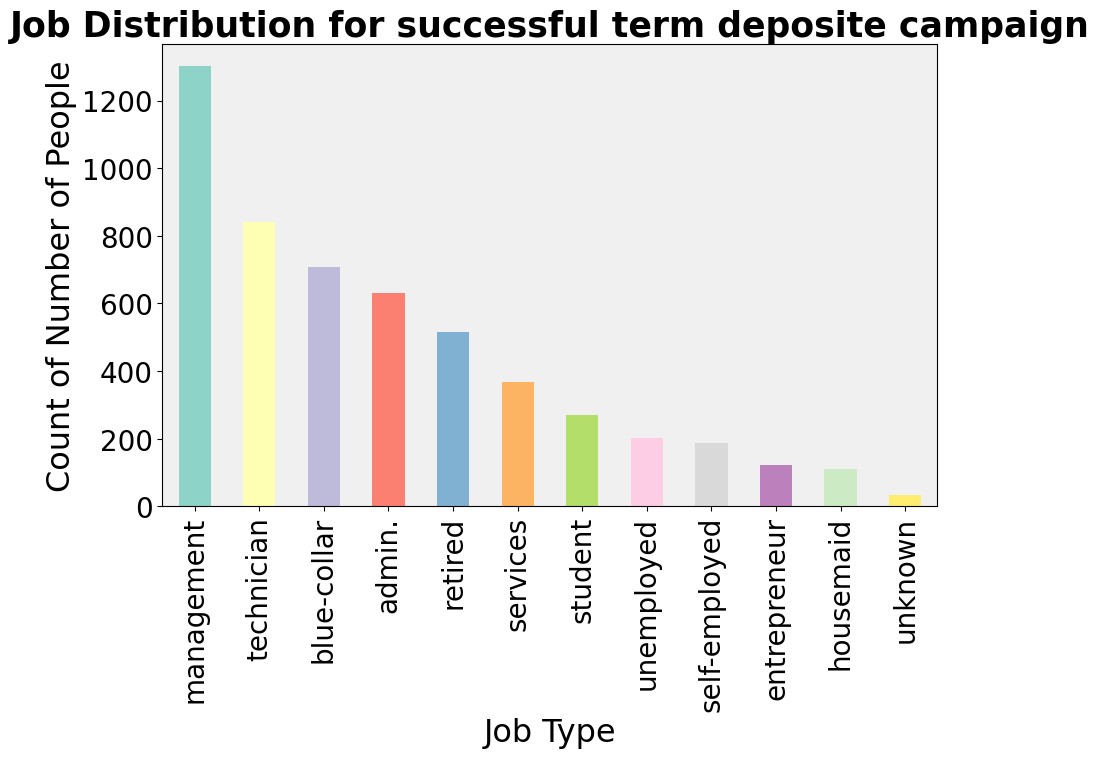
**Introduction:**

This data is related with direct marketing campaign by a Portuguese Banking Institution. The Goal of the Campaign was to classify weather the client will subscribe a term deposit or not. This dataset contains 16 features which affect the Term Deposit Classification. The dataset contains 768587 Data with 0 null values in any of the columns. Our Aim is to find when the Campaign is Successful which is when the client aggress to make a Term Deposit and try to find the factors affecting it .

**Data Analysis:**

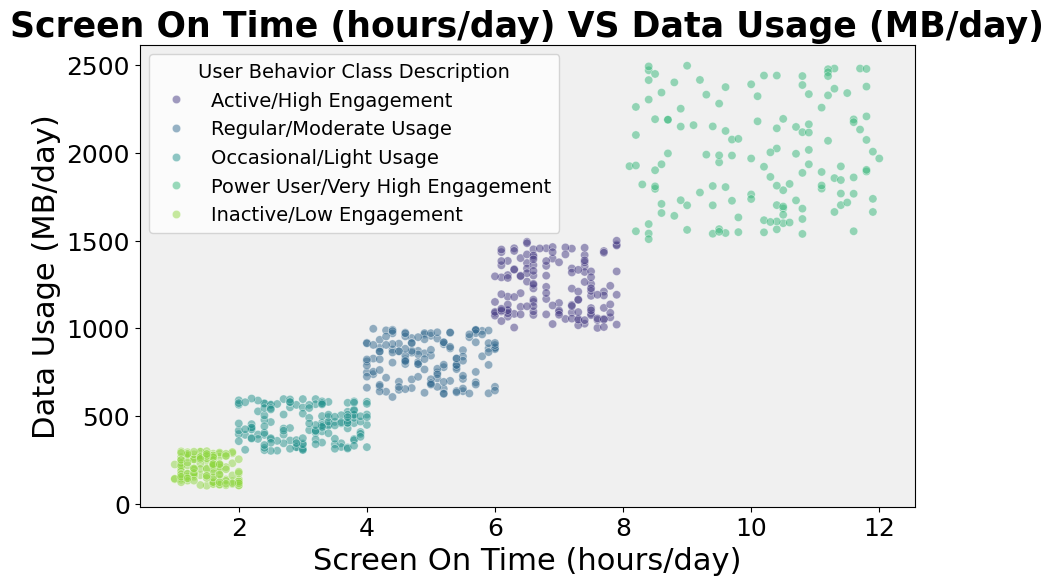
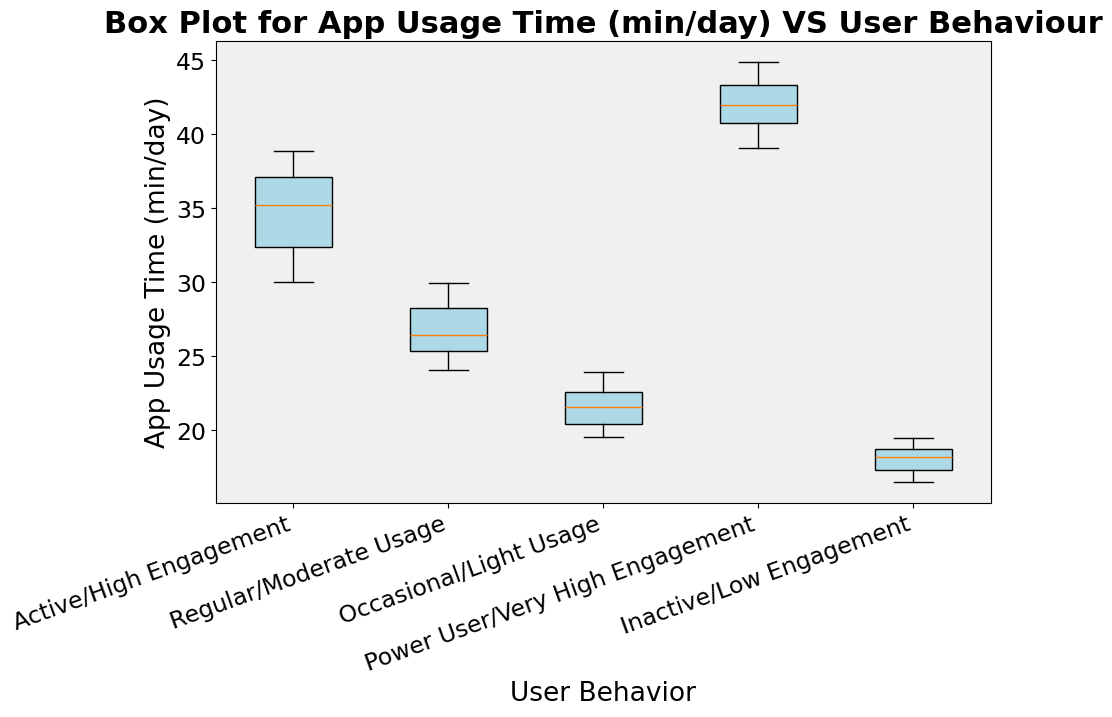
1. **Data Cleansing and Modification**:

Some of the columns like Day, Month, Duration were just representation of data provided in the pday column in different units so they were dropped. The average age of a client in the campaign aggreging for term deposit is 41 with the minimum age being 18 who is a Student by Job Type. There are 11 unique job types of the clients in the campaign. We would now mainly focus our analysis on the sole condition that the Campaign was successful to narrow down our Research.

1. **Bar Plot**

Based on the Bar plot we can say that more than half of the clients of the Successful Campaign are in Management sector followed by Technicians with a count almost 2/3rd of Management Sector. A additional observation can be made more than 88.2 percent of clients in self-employed, entrepreneur and housemaid sector are unwilling to have a term deposit as it makes sense because of the volatile nature of their sector .

1. **Scatter Plot**:

From Fig -1 we see that the Screen on Time of the device has a positive correlation with the Number of apps installed, the same we found before in the pascal correlation. In both Fig -1 and 2, the user distribution is almost uniform with a variation of +-10. And in addition to that, we can also say that the user behaviour has the highest engagement where the screen on time is large and so is the number of apps, installed and lowest when the values are less. From Fig-2 we can tell that the Screen on Time and Data Usage has a positive relation with each other, and we can derive that the User with the Highest Engagement (Power User) has more Data Usage and with Low Engagement (Inactive User) uses less.

**Fig -1**

**Fig -2**

1. **Box Plot:**

As we can see, the average app usage time for a Power User is greater than 40 minutes per day and for an inactive user, it is less than 20 minutes per day. In addition to it, we can also interpret that Power Users have a larger variation in the app usage time than other users. So, as we know from pascal correlation, we can say that if a User has a High App Usage Time, it would have a High Screen On Time also.

**Conclusion:**

After analysis we have observed ***Power Users/Very High Engagement*** have *High Screen On Time* and Significantly larger *Number of Apps* installed, have *High App Usage Time, High Data Usage*, and from pascal correlation we can say that the user would have *Higher Battery Drain Rate* with a higher probability of the device’s *OS* being an *Android Device*. For ***a Low Engagement / Inactive user*** we can say that the user’s *Screen On Time, Battery Drain Rate, App Usage Time, Data Usage* would be *less,* and the *Number of Apps* installed would also be less compared to other users. And there is a high probability of the user's device’s OS being Android.

**References**

*matplotlib.pyplot.bar*. (2024, 10 1). Retrieved from matplotlib.org: https://matplotlib.org/stable/api/\_as\_gen/matplotlib.pyplot.bar.html

*Mobile Device Usage and User Behavior Dataset*. (2024, 10 28). Retrieved from Kraggle: https://www.kaggle.com/datasets/valakhorasani/mobile-device-usage-and-user-behavior-dataset

Raoniar, R. (2024, 10 28). *Generate Numerical Correlation and Nominal Association Plots using Python*. Retrieved from medium.com: https://medium.com/the-researchers-guide/generate-numerical-correlation-and-nominal-association-plots-using-python-c8548aa4a663