**Introduction**

Under COVID-19 situation, many airline companies are facing a hard time in running the business. Being in a plane would be a high-risk activity because of the confined space. Also, many countries have shut down the border so as to minimize the way of spreading the virus. Therefore, people are not travelling overseas now and the revenue of the airlines companies drops significantly. In this project, the significance of factors which lead to customer satisfaction of the flight experience is evaluated. It is targeted to provide the insight to the airline companies for preparing the important criteria now, and provide better service to the passengers after the COVID-19 outbreak.

**Data**

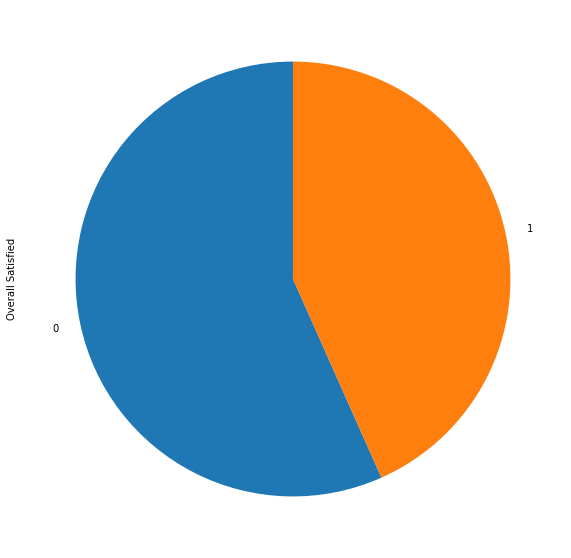
The dataset is found in Kaggle. There are about 104,000 rows in the dataset. Also, there are 24 columns in the dataset, including the background information of the passengers, the satisfaction level on different aspects, and the overall satisfaction level. Model of regression analysis would be built to evaluate the correlation between specific aspects of satisfaction level and the overall level.

**Data Preparation**

The data would be cleaned so that only used column would be left. As ID is irrelevant, the ID column would be dropped. Also, for the ease of analysis, overall satisfaction level will be changed into dummy variable, which "satisfied" would become True and "neutral or dissatisfied" would become False.

**Data Evaluation**

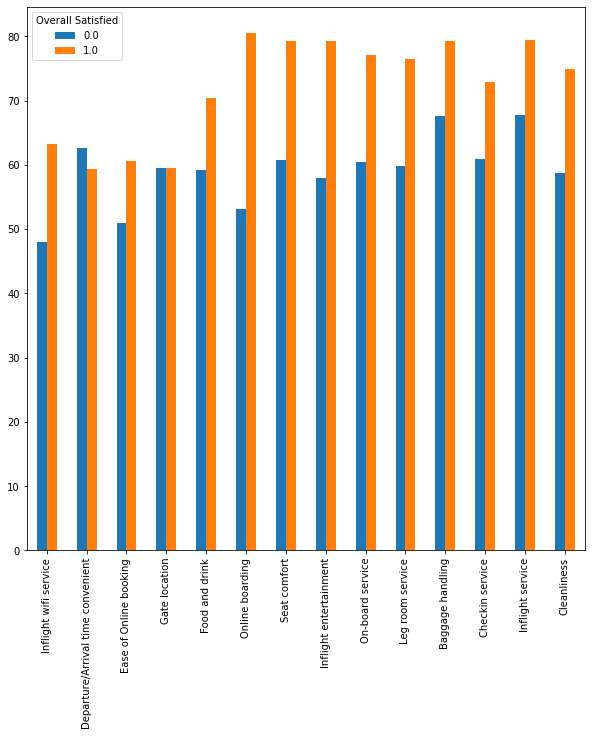
The following pie chart is the distribution of data regarding the overall satisfaction level.



From the result shown, this is a quite balanced labeled dataset.

**Single Factor Analysis**

Factors would be evaluated one by one to evaluate the significance to the overall satisfaction level. To have a better look, the satisfaction levels of the above factors would be multiplied by 20, so that the full satisfaction level would be 100. Also, visualization by bar chart can be performed so as to have an easier understanding of the data.



It is believed that the higher the satisfaction level in a single factor, the more chance that the overall satisfaction level is satisfied. From the above graph, the factors of online booking, inflight entertainment and seat comfort would greatly affect the overall satisfaction level, as the mean of satisfaction level which the group of people is overall satisfied is evidently greater than that of others. However, the factors of departure / arrival time convenient and gate location has the opposite result. For the factor of gate location, people who have the different overall satisfaction level have nearly the same satisfaction level in the gate location. For the factor of departure / arrival time convenient, people who are overally unsatisfied has a higher satidfaction level in the factor of departure / arrival time convenient than that of others. Therefore, this may illustrate that the factor of departure / arrival time convenient and gate location does not really affect the overall satisfaction level.

**Multiple Factors Analysis - Logistic Regression**

As the target is a binary variable, logistic regression is used to develop to use the factors to predict the overall satisfaction level. X would be factorsControllable[["Inflight wifi service", "Departure/Arrival time convenient", "Ease of Online booking", "Gate location", "Food and drink", "Online boarding", "Seat comfort","Inflight entertainment","On-board service","Leg room service","Baggage handling","Checkin service","Inflight service","Cleanliness"]] and Y would be factorsControllable["Overall Satisfied"].

After the fitting of the mode, the intercept is -7.62406913. The coefficients are 0.30472856, -0.29527595, -0.13290415, 0.12406888, -0.07312879, 0.82642633, 0.15850747, 0.29444351, 0.27612198, 0.35311107, 0.05478284, 0.21945884, 0.01345976, 0.03940116.

Therefore, the model built would be Overall Satisfied = -7.62406913 + 0.30472856(Inflight wifi service) - 0.29527595(Departure/Arrival time convenient) - 0.13290415(Ease of Online booking) + 0.12406888(Gate location) -0.07312879(Food and drink) + 0.82642633(Online boarding) + 0.15850747(Seat comfort) + 0.29444351(Inflight entertainment) + 0.27612198(On-board service) + 0.35311107(Leg room service) + 0.05478284(Baggage handling) + 0.21945884(Checkin service) + 0.01345976(Inflight service) + 0.03940116(Cleanliness).

From the above results, the sixth factor is the most significant factor affecting the overall satisfaction level, which is online boarding. The second most significant factor is the tenth factor which is the leg room service. The third most significant factor is the first factor which is the inflight wifi service.

**Conclusion**

From the above results generated, it is suggested the airline company should focus on online boarding service, leg room service, inflight wifi service, inflight entertainment and seat comfort. For example, they can upgrade the interface of online boarding service to be more user-friendly, update the inflight entertainment so that new movies and TV shows can be watched in the plane, enhance the coverage of wifi service in the plane, etc. By focusing at these factors, the overall satisfaction level of the passengers are believed to be increased efficiently.