1. What is the problem that you want to tackle?

   Bike-sharing is a service where bicycles are made available for shared use to individuals on a short-term basis. Generally, this business model has several different settings: such as docking station models where bicycles can only be borrowed and returned from designated docking stations, geo-fencing which unables bike hires to be ended outside of a virtual fence or free-floated bikes which allows the bikes to be dropped off at any location within a city's boundaries.

   In Singapore, the bike sharing industry is mainly dominated by dockless bikes offered by several private companies. The local bike sharing business starts in early 2017 when Obike rolled out its bikes to the public, followed by Ofo and Mobike both are established players in the bike-sharing market with a huge presence in China.

   After surveying different bicycle areas located in Singapore, our team has realised that there is an uneven spread of shared bicycles across different the stations in Singapore. What this means is that at some stations, we observe that that there are more parked bicycles than the number of docks allocated to that particular. On the other hand, some of the stations are totally barren and there are no bikes parked at all. Hence, we aim to create a program that will reflect the spread of bikes across different stations over the year of 2014. The bike sharing companies can then better distribute the number of bikes across different stations and this will also resolve the problem of shortages of bikes from the biker's point of view.

2. What is the current market practice

   i. Currently, shared bicycles work based on a QR code system. The QR codes are scanned through the App created by the company, after which the bike can then be used. Consumers can return the bike to any of the parking stations set up by the bike sharing companies where they then have to scan a QR code located at the parking zone to end their trip and park the bikes . These parking zones contain a geo-fencing system which only enables bikers to park and lock the bikes when the bikes are within the virtual perimeter (size of the parking zones) assigned by the biking companies.

   ii. All of the stations placed across the country accommodates a similar number of bikes, 5 at the small zones, 8 at the medium zones and 10 at large zones, which are normally only seen near MRT Stations or Bus stations. The current practices are comprehensible as MRT Stations or Bus Stations are locations with the most human flow and this is directly proportional to the rate of bikes rented out.

However, some of the locations with small or bigger zones also surprisingly, see a high rate of bikes being rented out.

3. How are you going to tackle the problem? What data to use and where to get them?
    i. Initially, our group intended to use data that are related to the local bike sharing companies that are in Singapore, such as OFO, Mobike and SGBike, since they are Singapore-based and hence, are more likely to be relevant for the purposes of this project. However, in the course of this project, taking into consideration the availability of these datasets and that most of the bicycle sharing companies such as OFO, Mobike and oBike have either exited or are planning to exit the local market, we have decided to use an alternative source of bicycle sharing data from San Francisco Bay Area. The final dataset we have chosen to use can be obtained from this link: https://www.kaggle.com/benhamner/sf-bay-area-bike-share.

    ii. Primarily, the dataset has been split into 4 different csv files containing the station, status, weather and trips data. In order to facilitate the process of data analysis, data cleaning was performed after the acquisition of data, by merging Stations, Status, and Weather into a single dataframe. From the combined dataframe, data was aggregated on an hourly average basis for us to understand the hourly trends on the availability of the bicycles and docks data for each station. In addition, we can also draw insights on the usage rate from these two metrics.

    iii. Before further analysis on weather effects on bicycle usage, a correlation matrix was generated for the three selected variables of analysis: Mean Temperature, Mean Humidity, and Cloud Cover. From the matrix, we can assume that there is no interaction between all the independent variables since their correlation coefficients do not exceed the threshold of 0.8, hence indicating no multicollinearity. Subsequently, further ANOVA and multiple linear regression analysis were then performed on the cleaned data and we found out that there was a correlation between the Usage Rate and all the independent variables of mean temperature, mean humidity and cloud cover. Since the p-value of ANOVA analysis falls below the statistical significance level of 0.05, we can conclude that there all 3 weather variables of analysis are good indicators on usage of bicycles. Using this understanding, we can help consumers predict with greater confidence the availability of bicycles and docks at each location for their bicycle sharing service.

4.

5. What are the pain points of current market practice to consumers? [HJ][bryon]

   Unlike other cities like London and Taipei, the local bike sharing business used to follow a free-floating system where bike users can park the bikes anywhere without penalties. Hence, such practice has resulted in issues like indiscriminate parking by bike users. To tackle this problem, the Land Transport Authority (LTA) has announced new regulations for dockless shared bike operators as well as the users. As reported by the Straits Times in September 2018, LTA will be installing Quick Response (QR) codes at public bicycle parking places to ensure that bikes are parked correctly. Furthermore, errant users will be charged $5 by licensed operators each time they park these bicycles indiscriminately.
   It seems to a good measure to stop indiscriminate parking of bikes since the new parking places will serve the same function as a docking station.

   However, given the limited space of a parking station, it may cause other issues as well. we foresee that this new law that was passed may cause problem in bicycle overcrowding at destinations that are popular end points, but less popular start points, whereas areas that are popular start points, but less popular end points might faces issues with bicycle shortages throughout the day. Besides, timing may also cause bike shortage or excess problems at various stations. For example, at places like the CBD area and during peak hours, the demand for renting a bike is likely to be high. Thus, causing bike shortages at stations in these area and customers may not be able to rent a bike at the station, vise versa. On the other hand, under the new regulation, bike users must park their bikes at designated QR code areas if they do not want to be fined. However, given the limited space at each station, there may be overcrowding problem as well where bike users have no choice but to park the bikes outside the designated area. This will cause another form of indiscriminate parking. Therefore, allowing customers to know the number of bikes available at the stations and to visualise the demand of bikes at certain stations under certain time intervals are essential in helping to deal with such problems.

6. How are you going to tackle the problem?
   i. What data to use and where to get them? [KJ]
      i. Bicycle trips data from major local bike sharing companies such as OFO, MoBike, SGBike, etc. (whichever is available)
      ii. In the event that local bicycle trips data is lacking, we can also consider taking similar data for bicycle sharing services from the other countries as references to better understand the different variables behind the popularity of start and end points of bicycle trips made. Ultimately, this

will allow us to better predict the demand requirements of bicycles at various bicycle hotspots at different timings throughout the day. (kaggle.com is another source of bicycle sharing data that we can consider)

ii. What kinds of data visualization to develop? [HJ][bryon]

i. Basic statistical charts to study user behaviour and pattern

1. Scatter plot for 2 numerical variables
2. Histogram for 1 categorical variable

ii. Spatial plots to visualise the demand and supply of bikes at several parking stations.
We have plotted

7. Strategies employed after analysis

i. <u>Dynamic pricing</u>

By implementing dynamic pricing, the cost of renting a shared bike will be dependent on the demand (Rate of bikes getting rented) and the availability of bikes at both the starting station and ending station. Getting a bike from a high demand, low availability station will cost users more as compared to getting one from a low demand, high availability one. Similarly, returning the bike to a high demand, low availability station will cost users less as compared to returning one to a low demand, high availability one. By doing so, users will be encouraged to pick up their bikes from stations with greater number of bikes and return them at stations with lesser number of bikes. This will help to resolve the issue of an uneven distribution of bikes and also the issue of shortage of bikes at selected stations.

ii. <u>Redistribution of supply of bikes</u>

As seen from the ggmap above, some areas observe an excess in the number of bikes available and likewise, other areas observe a shortage in the number of bikes available. One easy way to resolve the uneven distribution of bikes is to redistribute the supply of bikes according to the data. Bikes at stations with an excess number of bikes can be relocated to stations that observe a shortage in the number of bikes available. This strategy is also relatively easier to implement in Singapore as the bikes are not required to put park in docks at the stations, and would only need to be parked in the designated geo-zone.