* Introduction where you discuss the business problem and who would be interested in this project.

Mayhan is a property dealing company in Kabul province. They want to analyze districts in Kabul city based on their shared characteristics so they could provide a clear picture to their customers about how a particular district in Kabul may match their taste based on the available venues in each district.

* Data where you describe the data that will be used to solve the problem and the source of the data.

we used the four-square API to extract the data of available venues with respect to a particular radius from the points of interest (depending on area of each district). Then we extracted top ten venues for each district based on which we did our analysis. (more information about the data is available in the notebook)

* Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

Here we used k-means clustering machine learning algorithm with total number of 3 clusters. We used this unsupervised machine learning algorithm because of the following reasons.

1. We used clustering because we didn’t know too much about the characteristics of the data beforehand to use any classification algorithm and as Kabul is an expanding city this algorithm can be used on the updated data and k values to gain new insights and groups of neighborhoods.
2. First we used value 5 for k (number of clusters) after looking through data of each cluster we observed that some clusters were very similar than we changed the number of clusters to 3 to achieve the optimal number of viable number of clusters.

* Results section where you discuss the results.

The total number of 3 clusters were yielded after the analysis after observing each cluster closely it was concluded that:

1. The first category of districts (Cluster label ‘0’) are densely populated by having lots of coffee shops, restaurants in a relatively small area. And may fit the taste of those who loves very densely populated areas.
2. The second category of districts (Cluster label ‘1’) are not so densely populated places with having historical sites (palaces) and agricultural sites (farms) people who love quite places will love living at these places.
3. The third category of districts (Cluster label ‘2’) are suburb type places where they are neither very populated nor less populated and having a few restaurants in the vicinity. Its said that rents in such places are lower so can be a good option for people on low budget.

* Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

As Kabul is drastically expanding city things may change very rapidly so these reselts may not be viable for use in the long. using this algorithm by changing the k value and updating data can easily give us the updated picture of the city from which we can exclude new results. So it is recommended that this analysis should be done in each quarter. Using the historical results extracted from this analysis we can also have a trend of common venues in each area that can be a got informative tool for investors in Kabul city

* Conclusion section where you conclude the report.

Majority of the districts in Kabul city are densely populated with having lots of venues around which results in traffic only 2 districts were categorized as quite and low populated cities with having historical sites. This study is based on the most recent data and as with the expansion of the city the data will also change so it will be very important to carry on same analysis on the updated data to receive new results.