**Project Idea:**

Airbnb is one of the most used methods to find a place to stay in most of the major cities in the world. It has dramatically changed the traditional way to find accommodations for short periods of stay and even longer ones. I would mainly like to concentrate on the following questions being raised:

* What is the seasonal pattern of Airbnb in Seattle?
* What kinds of Airbnb homes are popular?
* Prediction of the price.

The outcome of this project would help businesses narrow down their target marketing customer base towards the most likely prospective customers and this project would also help the business to understand how various attributes affect the client to accept or reject the Airbnb home.

**Data Storage Platform:**

The data that I would like to use for this project is from http://insideairbnb.com/get-the-data.html From this I would like to use the following csv data sets.

* Calendar.csv
* Listings.csv
* Reviews.csv

To store the data I prefer to store it in the Big query.

**Data Transformation:**

Data is available in 2 forms, as streaming data which is streamed via pub/sub topic, and the data in form of CSV files which are injected into cloud storage using gsutil.

I prefer to use Data fusion for the ETL process and build pipelines to store it in cloud storage after the ETL process and store that in form of tables in Big Query for better visualization and analysis.

**Description of the dataset:** Listing.csv contains more than 15000 records and 92 attributes. Calendar.csv contains nearly 1300000 records and 4 attributes Reviews.csv contains more than 90000 records and 6 attributes. This is a categorical dataset. I plan to inject the dataset to a big query via pub/sub topic. After storing the data, the data is cleaned based on the requirements.

**Visualization and Analytics:**

For the first two questions, I would like to use visualization and EDA techniques to answer the questions. For the last question I plan to use Linear Regression and Ridge and Lasso regression techniques. Mean\_squared\_error and r2\_score would be used for the model evaluation.

The stored data in Big Query is sent to Data Studio for view of the visualizations.

The stored data in Big Query is also sent to Cloud Datalab to explore, visualize and transform the data using python to predict the prices of the Airbnb home. I prefer Cloud Datalab to build the model using Linear, Ridge, and Lasso Regression.