Throughout this course and project we learned how to effectively create a data pipeline from several data sets, visualize the database with the ingested data using ERDs, implement and create metrics using SQL in BigQuery in order to visualize them, and analyze the data to infer valuable insights. With this came many challenges along the way, but this only allowed us to improve our debugging and analytical skills. Our main focus for this project was to dig deeper into the given Airbnb data for the city of Austin with a data-driven analysis by exploring the revenue generated over time, the occupancy rate over time, and the amount of rentals available over time.

We initially setup our database using Google’s GCS, Postgres, and Github, which were all invaluable tools for this project. Github was extremely useful in allowing us to share our files, code, and other relevant work, allowing use to easily work in collaboration. We ingested the Airbnb data into the GCS, and later did the same with the Zillow dataset in order to draw comparisons between the two. We created a view called the “Revenue Crossover Point” which really let us gain some insight into the comparison between Zillow and Airbnb’s prices.

Once we had the database set up, we dived into some analyses by creating relevant metrics with SQL and visualizing them in BigQuery. The first metric we created and visualized was the revenue generated from the Airbnb rentals. From the visualization we inferred that while the average revenue has taken a slight dip over time, it still remains at a relatively high and constant rate. For the second metric, we addressed the number of Airbnb rentals available over time in the city of Austin. Since the city of Austin has been growing in popularity over the last few years, we expected that the number of rentals would raise over time, and that’s exactly what our results were. The surprising part of this though was that there was a span from about June 2016 to October 2016 where the number of rentals actually decreased. This could be due to the error we mentioned earlier about the data, where the was no data from October 16, 2016 to February 21, 2017. For our third, and last metric…