# Project Summary and Discussion

In this project we analyzed Australian Taxation data from 1999-2000 to 2016-17 income years. The data is based on number of individuals with interest in rental property by number of property interests and either net rent position, taxable income range, state/territory or age range.

The dataset is consists of 4 parts or sheets net rent position, taxable income range, state/territory or age range, so we analyzed them separately. Table 27A is based on net rent position, 27B on income range, 27C on state/territory and 27D on age range.

After cleaning the data we did an exploratory data analysis to understand behavior of the dataset. We plotted heat map to see if any data was missing. Then plotted correlation heat map to get idea how data is related to each other. Later we plotted pair plot, joint plot, count plot, box plot and lmplot to get better idea about the datasets. After analyzing data we used several algorithms like Linear Regression, ARMA, ARIMA, KNN to predict the future property interest and number of individuals. After building models we evaluated those using metrics like MSE, RMSE, MAE and Confusion Metrics. We did not used K-means algorithm though it was in our project goal, because our data is already defined in different categories so we don’t need to cluster it for further analysis.

**Net rent position (27A):** Table 27A we saw the number of individuals who has overall net rent loss and has single property interest are increasing over the years. But After 2010 it was getting flatter until 2015. We did not see any seasonality in that data set and data was stationary. Other than single property interest when it comes to multiple property interest (more than two) the number of individual those who has overall net loss their number was constant in almost all the years. The data was stationary non seasonal.Later we used Linear Regression, ARMA, ARIMA and KNN to predict number of individuals for 3 years in future.

**Taxable income range (27B):** In table 27B we saw those who has taxable income range below 18000 dollars their amount was not increasing over the time period 1999-2017. But when it comes to taxable income range 18000-50000 dollars the number of individual with single property interest was decreasing significantly over the time period. We did not found any seasonality on the data and data was stationary. The data was stationary non seasonal.Later we used Linear Regression, ARMA, ARIMA and KNN to predict number of individuals for 3 years in future.

**State territory (27C):** In table 27C we found that in state of ACT the number of individual who has single property interest was increasing significantly over the time period. But when it comes to multiple property interest the number was getting flatter over the years. The data was stationary non seasonal.Later we used Linear Regression, ARMA, ARIMA and KNN to predict number of individuals for 3 years in future.

**Age range (27D):** After analyzing table 27D we saw the number of individuals with single property interest and age range between 40-49 years was increasing significantly over the time period of 1999-2017 fiscal years. The number was increasing linearly with slope of 1. The number of individual with single property interest and age range 30-39 was also increasing. The data was non stationary non seasonal.Later we used Linear Regression, ARMA, ARIMA and KNN to predict number of individuals for 3 years in future.