# Team 8: Roujing Chen, Shelley Shi, Wentao Shi, Shouzhi Wan

## *Data Cleaning*

After loading the data into R-Studio, we find that many variables' representation are repetitive and redundant, for instance, "job class" and "job class title", "MOU" and "MOU title." As well as what we believe to be irrelevant information to our prediction like "job class link" and "benefit costs." Thus, we take a step back and clean them out of the dataset.

In addition, Model built on highly correlated data would result in biased prediction and unbalanced weights. In the original dataset, "total payment" represents the sum of the quarterly payments as well as the total of base pay, other pay and overtime pay. As a result, we performed addition data cleaning and truncated the dataset to only relevant variables.

The ones we decide to keep as variables are "Year, Record Number, Employment Type, Q1 Payments, Q2 Payments, Base Pay, Overtime Pay, Other Pay(explorer), MOU, Pay Grade, and Benefits Plan."

## *Model building*

We decided to follow the method of using previous year's data to predict current year's Q1 and Q2 payroll given we have limited information in the year of 2018. The model we built explores the connection between these variables. We also examined the trend of payroll growth for the same quarter of different years and the line chart of Q1 payments changing along the years indicates that the overall payroll is gradually increasing. Based on careful consideration, we find



Fig. 1 Trend of Total Payment

**Multi Linear Regression** to be the optimal model for prediction of Q1 and Q2 payroll in 2018.

The model is trained on 2 consecutive years of selected data (ex. we use the 2013 data to predict 2014 payroll and so on). Eventually, we use the trained model to predict 2018's payroll with the trained model. Below is the validation result of our model.
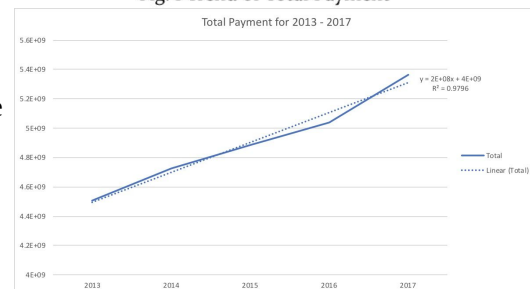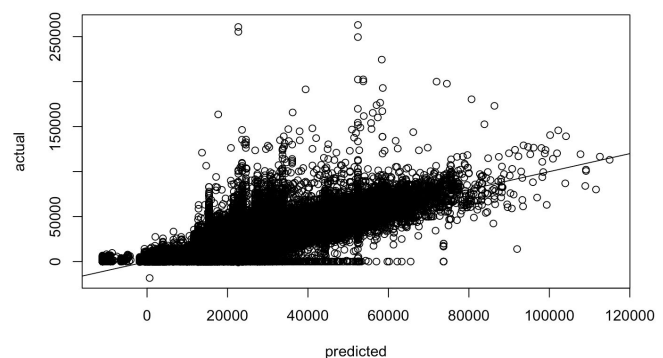
Fig. 2 Validation Result



Fig. 3 Multi Linear Regression Model

After validation, the model indicates most correlated factors that influence the prediction. The plot below shows the difference between the value of actual Q1 payments and the predicted value of Q1 payments from our multi linear regression model from year 2013 to 2017.

## *Optimization & Suggestions*

Regarding the optimization plan, we highly recommend the City of Los Angeles to review the base payment for public accountability and overtime payment policy for the Fire department. Because, base on the record, the average base payment seems to be fine except for an unusual spike within public accountability. In terms of overtime pay, fire department skyrocketed the record.

Fig. 4 Average Overtime Pay among Departments
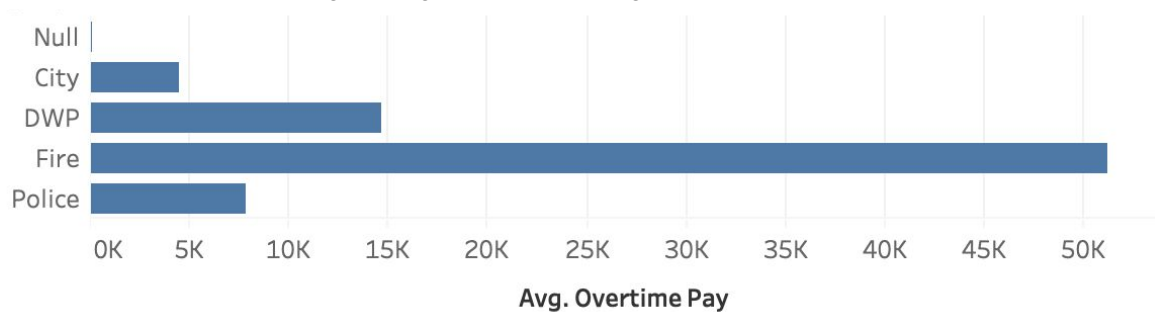


After digging into the overtime payroll, we observed top 5 job titles for average overtime pay are from the fire department(pilots). We understand given the fact that wildfire is a frequent disaster in California, the fire department can be extremely busy over the year. We also sought information from other sources which all express surprises regarding high total wage for firefighters. Some analyses suggest that firefighters at least are working at their maximum abilities. We understand firefighters takes serious risk in their daily job, but we believe the composition of overtime and base pay is unhealthy. The city officials should consider allocating more resources into fire prevention – especially in mountain fire, as well as hire more firefighters to even out the balance of base pay and overtime pay. This will also benefit firefighters, since less overtime will give them healthier work-life balance and meanwhile lowering the risk of injuries.