Statistical Analysis of San Francisco public employees’ compensation

Anthony Philpott, Sweta Kumari, Tanapat Klomjit

Golden Gate University

Statistical Analysis of San Francisco public employees’ compensation

Table of Content:

1. Introduction
2. Data collection and Description
3. Statistical Analysis with Interpretation
4. Conclusion
5. Reference

# Introduction

# San Francisco City is the 13th most populous city (NLC, n.d.) in the United States of America with a population of 884,363. With the increasing number of people, employers, technology in the city, the cost of living has increased and become 80% more than the public average (Payscale, n.d.). There could be several factors which have increased the cost of living in the city; analyzing and forecasting the compensation of employee will help to get the idea about the next year's compensation of individuals. San Francisco Controller’s office captures data of public employees of the city of San Francisco that can help to gain insight into the different benefits and salaries considered for employee compensation. Due to continuous advancement in the technology, Information System Business Analyst job has evolved from past decades among all jobs. We have filtered the job and chose to predict the compensation of Information System Business Analyst job among 1300 jobs. To understand the compensation of previous years and forecast the same for the next 10 years, descriptive analytics, multiple regression analysis, and Facebook prophet procedure have been performed on the employee data set. We analyzed the variability and distribution of the total compensation and probability of getting higher compensation for IS Business analyst job. We inspected the data set to find about which department gets highest compensation and how salary is increased per quarter and years for the department. For this, we performed multiple regression for time series data to predict salary and tried dropping/adding different predictors till the model was significant. We used Facebook prophet procedure to predict and forecast the monthly average compensation for public employee.

# Data collection

To perform the statistical analysis, we collected data from the Kaggle website: <https://www.kaggle.com/san-francisco/sf-employee-compensation> . Due to a wide variety of datasets from various industries and employers, we chose Kaggle among all open sources dataset websites. Initially, we downloaded the full dataset for the city of San Francisco employee for the year 2013 to 2019 with the calendar and fiscal years. This data was updated on a bi-weekly basis in the database of the San Francisco Controller's Office. We started inspecting the data for our best interest which narrowed down our analyzation in terms of a job. Since, pursuing a master’s degree in business analytics (we all have major as business analytics) and, frequent fluctuation in the compensation in this field has led me to choose a job as an Information System Business Analyst among all jobs available in the original dataset. We had to further refine our dataset and ignore all the null values from the original dataset.

**Goals and Variables**

We are focusing on Information System Business analyst job. With statistical analysis, we aim:

a. To understand the variability and distribution of the total compensation.

b. To find probability of getting high and low compensation.

c. To know the department name which get highest compensation.

d. how the salary is increased over the quarter and years in the department which get highest compensation?

e. Forecast the average monthly compensation for next 10 years.

For time series multiple regression, we operated the total compensation as a dependent variable for POL Police department. Other variables include total salary, Department, overtime, other salaries, total salary, retirement, health/dental, other benefits, total benefits. We have removed Organization Group, Organization Group code, Department Code, Union Code, Union, Job Family Code, Job Family, Job Code, to simplify the analysis. After setting the job as Information Systems Business Analyst, the modified data set with a column header and description is shown below:

Table1: Data description after modification

|  |  |
| --- | --- |
| **Column Header** | **Description** |
| Year Type | Fiscal (July through June) |
| Year | An accounting period of 12 months. |
| Department | Departments are the primary organizational unit used by the City and County of San Francisco. Examples include Recreation and Parks, Public Works. |
| Job | Jobs are defined by the Human Resources classification unit. Here we are treating for Information systems business analyst job |
| Employee Identifier | Each distinct number in the “Employee Identifier” column represents one employee |
| Salaries | Normal salaries paid to permanent or temporary City employees. |
| Overtime | Amounts paid to City employees working in excess of 40 hours per week. |
| Other Salaries | Various irregular payments made to City employees including premium pay, incentive pay, or other one-time payments. |
| Total Salary | The sum of all salaries |
| Retirement | City paid amounts to employee retirement plans. |
| Health/Dental | Premiums paid to health and dental insurance plans for City employees. |
| Other Benefits | Other benefits paid on behalf of employees not included in the above categories, such as Social Security (FICA and Medicare) contributions. |
| Total Benefits | The sum of all benefits. |
| Total Compensation | Total Salary + Total Benefits |

# Statistical analysis

**Descriptive Analytics:**

**Box and Whisker plot**: We started our analysis with box and whisker plot which give us an idea about the Total compensation for IS Business analyst in the different quartile with maximum and minimum amounts. The box plot shows variability, the shape of the distribution and central value of the total compensation. We removed outliers to calculate all three quartile Q1, Q2, and Q3. Inter Quartile Range, upper fence and lower fence are determined to detect unusual data points in the data set.

Table2: Quartiles calculation Table3: Fences calculation

|  |  |
| --- | --- |
| **5 summary rules** | |
| Q0(min) | 664.29 |
| Q1 | 70025.56 |
| Q2(median) | 120822.975 |
| Q3 | 135022.8025 |
| Q4(max) | 163169.95 |

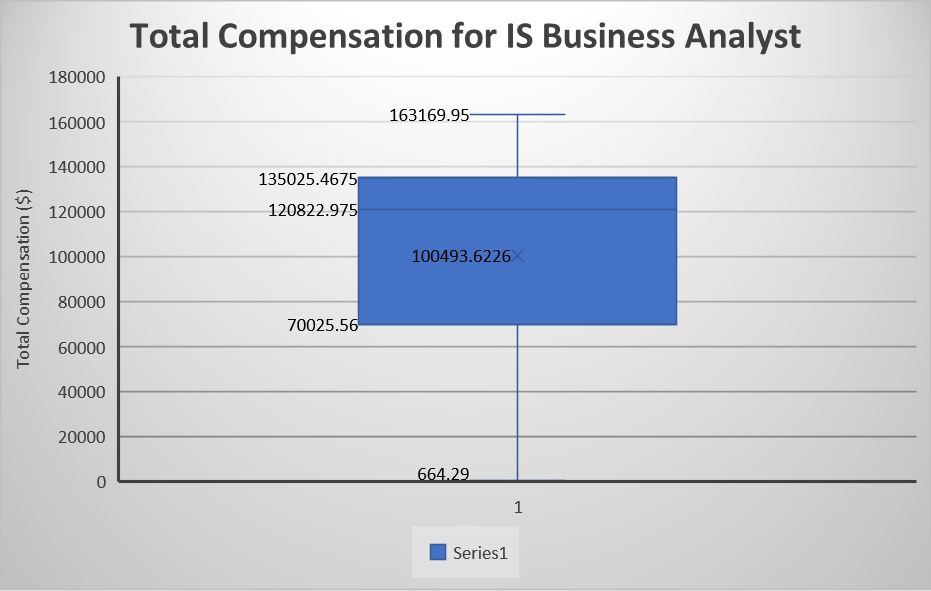
|  |  |  |
| --- | --- | --- |
| **1.5 IQR calculation** | | |
| IQR | Q3-Q1 | 64997.24 |
| Lower fence | Q1 - 1.5IQR | -27470.3 |
| Upper fence | Q3 + 1.5IQR | 232518.7 |

Lower fence= -27470.3 (or Q0, whichever is bigger)

Upper fence= 232518.7 (or Q4, whichever is smaller)

Below is the box and whisker plot:

Graph1: Total compensation for IS Business Analyst



**Interpretation of box and whisker plot**: From graph1, we illustrate that the distribution of the compensation is left-skewed where the median (120823) is greater than mean (100493.6). It has a longer tail on the left side and most of the data is distributed on the right side. The median is close to the third quartile. The maximum compensation (Q0) is $1,63,170 while the minimum compensation (Q4) is $664.3. First, 25% (Q1-Q0) of the employee has compensation ranges from $664.3 to $70,025.56 while the Second 25% (Q2-Q1) of the employee has ranged from $70,025.56 to $1,20,823. Third 25% (Q3-Q2) of the employee has compensation ranges from $1,20,823 to $1,35,022.8 while fourth 25% (Q4-Q3) has ranged from $135022.8 to $1,63,170. Most of the employee has compensation on the higher side due to left-skewed distribution.

**Probability for compensation:** After finding the variability in the total compensation, we did find the probability for the IS Business Analyst job. Below table shows the different compensation group with the number of employees who has a job as IS Business analysts.

Table4: Compensation amount with number of employees

|  |  |
| --- | --- |
| **Total Compensation Amount** | **# IS Business Analysts** |
| Under $50000 | 161 |
| $50000-$79999 | 51 |
| $80000-$99999 | 26 |
| More than $100000 | 370 |
| Grand Total | 608 |

From table4, we illustrate that the maximum number of employees is earning higher compensation i.e, more than $1,00,000

Probability that an IS Business Analyst will earn yearly compensation >= $80,000

P1=P (IS Business Analyst Compensation >= $80,000) = (26 + 237) / 474

P1=65.13%

P2=Probability that an IS Business Analyst will earn yearly compensation < $50,000

P (IS Business Analyst Compensation < $50,000) = (160) / 474

P2=26.48%

**Interpretation of Probability:** The probability that an IS Business Analyst will earn yearly compensation >= $80,000 is 65.13% while Probability that an IS Business Analyst will earn yearly compensation < $50,000 is 26.48%.

**Average compensation with respect to department:** We calculated the average yearly compensation as per departments to know which department get highest compensation for job IS Business Analyst.

Graph2: Average of Total Compensation for IS Business Analysts in SF

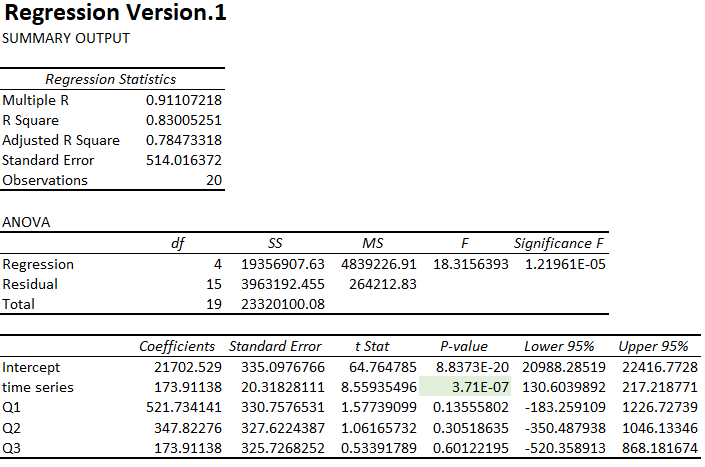
**Interpretation of** **Average compensation with respect to department:** From graph3, we illustrate that the highest compensation paid to the employee who belong to POL Police department. The employee of HSS Health service system get the second highest compensation.

**Multiple Regression:** From above analysis, we got the result that the employee of POL Police department get highest compensation in San Francisco city. From Time series multiple regression, we analyzed out how salary has been increased for this department.

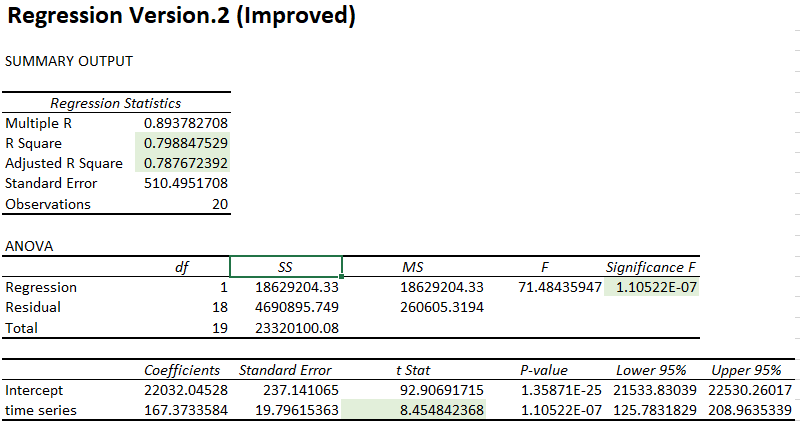
In data set, we had data for 2013 to 2017 which was used to plot below line graph (graph3).

Graph3: Quarterly total salary for IS Business Analyst in SF POL Police department 2013-2017

We ran multiple regression for time series data (Regression version1) including each quarters and time index.



The purpose of running first regression analysis is to determine significant predictors by looking at P-Value. As a result, only time series factor has a significant P-value. Thus, we will drop off insignificant value for the next improved version.



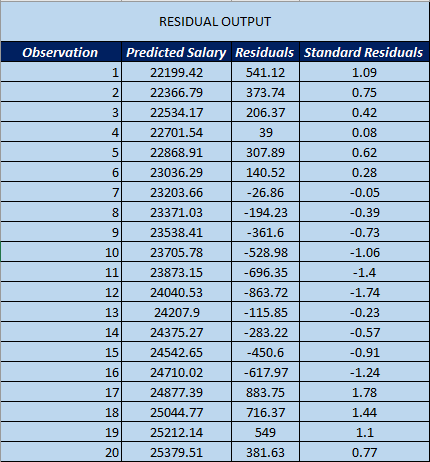
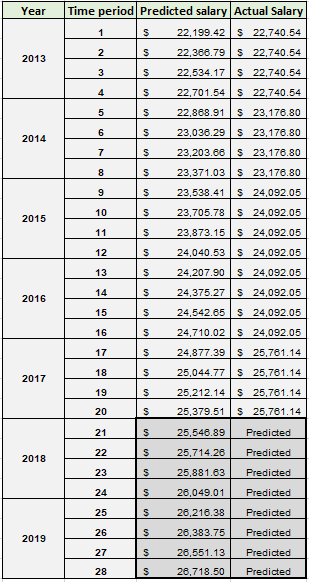
By analyzing at P-value in the second regression, the value is significant. R square and adjusted R square values are closer together. The model is improved after removed insignificant values. Thus, it will be used to build a final regression model equation.

Final Regression Equation:

Coefficient explanation:

Based on the regression equation, time series has a positive coefficient, so the total salary for IS business analyst in POL Police department is increasing $167.37 for every quarter pass. In another word, the trend of the total salary is increasing over time.

Residual Output Predicted salary for year 2018 and 2019



We forecasted the total salary for years 2018 and 2019 for POL Police department.

Graph4: Actual Total Salary vs Predicted Total salary

**Line graph of time series data for year 2013 to 2019:** We calculated the average of all attributes for Information Systems Business Analyst job in each year. From 2013 to 2019, we plotted a line graph for time series data for variables total salary, overtime, other salaries, total salary, retirement, health/dental, other benefits, total benefits, and total compensation. Before plotting the graph, we took an average of all attributes for each year from 2013 to 2019

Graph5: Line graph for time series data

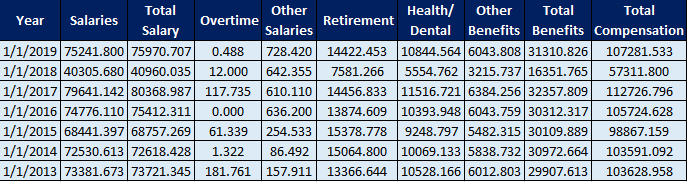
A screenshot of a cell phone

Description automatically generated

Below is the table shows the average values of different attributes for IS Business Analyst job for each year from 2013 to 2019. Total Compensation is the sum of all salary and benefits;

Total compensation = Total Salary + Total Benefits

Table5: Average yearly compensation with all attributes from year 2013 to 2019



**Interpretation of line graph of time series data:**

From the graph2 and table5, we see the fluctuation in different components. In 2018, a sudden drop in average salary, average retirement, and other salary have been observed. As total compensation is the sum of total salary and total benefits, the average total compensation is also dropped and shows the least value as $57,311.8. There is no continuous increment in the total compensation in last 6 years.

**Forecasting compensation using Facebook Prophet:**

We used prophet procedure to forecast the average compensation of IS Business Analyst monthly. Before proceeding with the procedure, we calculated the average monthly total compensation for the year 2013-2019.

Table6:



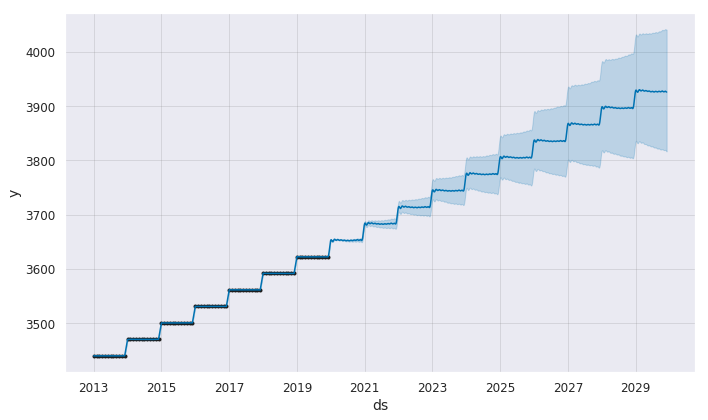
We uploaded the above monthly data set to python and imported Facebook Prophet. After importing, we used the function to fit and predict monthly average compensation. The notation is used as Yhat for predicted monthly average compensation for the next 10 years from 2020-2029. Other dimensions yhat\_lower and yhat\_upper included showing the fluctuation in the compensation within range. We used the multiplicative model for forecasting the data as follows:

Data = (Seasonal effect) x Trend x Cyclical x Residual

log(Data) = log(Seasonal effect x Trend x Cyclical x Residual)

Below graph shows the forecast for average monthly compensation for year 2020-2029 using data for years 2013-2019:

Graph6:



Below graph shows the seasonality trend for yearly and monthly basis:

Graph7:

A screenshot of a cell phone

Description automatically generated

From the graph, we illustrate that both seasonality and original data increase over time

Below is the result for average monthly compensation for IS Business Analyst job:

Table7: Forecasting the average monthly compensation for next 10 years

|  |  |
| --- | --- |
| **Year** | **Yhat Average** |
| 2020 | $3655.243 |
| 2021 | $3685.778 |
| 2022 | $3716.291 |
| 2023 | $3744.24 |
| 2024 | $3777.372 |
| 2025 | $3807.907 |
| 2026 | $3838.42 |
| 2027 | $3868.933 |
| 2028 | $3899.5 |
| 2029 | $3927.387 |

# Conclusion

1. Box and Whisker Plot: As per table2 and table3, we analyzed that the maximum compensation for IS Business analyst is $1,63,170 while minimum compensation is just $664.30 considering all departments. Majority of the employee in San Francisco has compensation on the higher side with average yearly compensation as $100493.6. 50% of the employee whose salary is on the higher side has ranged from $70,025.56 to $1,35,022.8.
2. Probability for compensation: The maximum number of employees in San Francisco (more than half of the employee) is earning higher compensation i.e, more than $1,00,000 (table4). The probability that an IS Business Analyst will earn yearly compensation >= $80,000 is high as 65.13% while Probability that an IS Business Analyst will earn yearly compensation < $50,000 is low as 26.48%.
3. Average compensation with respect to department: In San Francisco City, the public employee who belong to POL Police department get highest compensation among all departments.
4. Time series multiple regression: We predicted the total salary for POL Police department for years 2018 and 2019 which shows salary is increasing over the years.
5. Line Graph for Time Series: We analyzed the fluctuation in all components and found that there is a sudden drop in total compensation in San Francisco for a job IS Business Analyst. It was due to other components such as salary, overtime, retirement amounts which were also dropped.
6. Facebook Prophet: In San Francisco, the average monthly compensation for IS Business Analyst will increase over the years. In next year, monthly compensation would be $3655.243 while after 10 years, it would be $3927.4. We conclude that the IS Business Analyst job will have a continuous increment in compensation for the next decade.

References

Kaggle. (2019). *SF Employee Compensation From San Francisco Open Data*. Retrieved from

<https://www.kaggle.com/san-francisco/sf-employee-compensation>

Payscale. (2019)*. Cost of Living in San Francisco, California.* Retrieved from

https://www.payscale.com/cost-of-living-calculator/California-San-Francisco

NLC. (n.d.). *The 30 Most Populous Cities*. Retrieved from https://www.nlc.org/the-30-most-

populous-cities.