Objective

This project has the objective of using Tableau as a data visualization tool in order to create an interactive dashboard using the data set and creating data charts in order to get some useful insights.

We have made use of the visual analytics in the following manner:

- Reference lines and trend lines to visually evaluate the plots
- Forecasting to make predictions with a confidence level
- Statistical tools like average lines, median and Box Plots

Data Set

When a US company wants to hire someone from outside of the United States for a technical position, they have to file an application to the United States government to get a green card or visa for the foreign applicant. These applications allow the US government to track who is entering and leaving the country for work-related reasons, and ensure that immigrants are neither being taken advantage of nor causing adverse effects for U.S. workers. To ensure equity for US and non-US workers, companies have to state how much they are planning on paying the employee every time they submit a visa or green card application. They also have to state the average amount an employee with similar skills and background typically gets paid for the same position, a figure called "the prevailing wage." This publically available data provides a unique view into what types of salaries you might encounter for different data–related jobs in the US.

Source

The original data was compiled by the US Department of Labor's Office of Foreign Labor.

Overall Description

We assembled 167,278 records from the Labor Condition Applications (which include H1-B and other specialty visas) and the permanent resident applications from 2008 to 2015. The number of H1B and permanent visa applications considered in the dataset can be seen in the table below.

DATA OVERVIEW

Variable Description:

For each visa application in the dataset, the following information was prepared. Data was discarded or cleaned when necessary.

FIELD NAME IN DATASET	DESCRIPTION
Case Number	Unique identifier assigned to each application submitted for processing to the ETA National Processing Center.
Case Status	Status associated with the last significant event or decision in the case application. Valid values include "Certified," "Certified-Expired," "Denied," and "Withdrawn."
Case Received Date	Date the application was received by the ETA National Processing Center.
Decision Date	Date on which the last significant event or decision was recorded by the ETA National Processing Center.
Employer Name	Name of employer requesting certification.
Prevailing Wage Submitted	Prevailing wage for the job being offered.
Prevailing Wage Submitted Unit	Unit of Pay. Valid values include "Hourly (hr)," "Weekly (wk)," "Bi-Weekly (bi)," "Monthly (mth)," and "Yearly (yr)."
Paid Wage Submitted	Lower range of the wage offered to applicant
Paid Wage Submitted Unit	Unit of Pay. Valid values include "Hourly (hr)," "Weekly (wk)," "Bi-Weekly (bi)," "Monthly (mth)," and "Yearly (yr)."
Job Title	Payroll title of the job being offered, as entered on the

	application.
Education Level Required	**Minimum level of education required to adequately perform the duties of the job being offered.**
College Major Required	**Major field of study required based on the education requirement.**
Experience Required Y N	**Identifies whether experience in the job offered is a requirement (Y/N).**
Experience Required Num Months	**The number of months experience that are required for the job (if applicable).**
Country Of Citizenship	**Country of citizenship of the foreign worker being sponsored by the employer for employment in the United States.**
Prevailing Wage Soc Code	**Occupational code associated with the job being requested, as classified by the Standard Occupational Classification (SOC) System.**
Prevailing Wage Soc Title	**Name associated with the Prevailing Wage SOC Code.**
Work City	**City information of the foreign worker's intended area of employment.**
Work State	**State information of the foreign worker's intended area of employment.**
Work Postal Code	**Zip Code information of the foreign worker's intended area of employment.**
Full Time Position Y N	**A binary variable indicating whether the position is full-time (Y) or part-time (N).**
Visa Class	**Indicates the type of application submitted for processing. Options include "H-1B," "E-3 Australian," "H-1B1 Chile," "H-1B1 Singapore," "greencard."**
Prevailing Wage Per Year	**Prevailing wage per year, calculated for this project from the "Prevailing Wage Submitted" and "Prevailing Wage**

	Submitted Unit" fields.
Paid Wage Per Year	**Paid wage per year, calculated for this project from the "Paid Wage Submitted" and "Paid Wage Submitted Unit" fields.**
Job Title Subgroup	**Subgroup label made for this project course based on the "Job Title" field. Options include assistant professor, attorney, business analyst, data analyst, data scientist, management consultant, software engineer, teacher.**

ABOUT TABLEAU

- Tableau is groundbreaking data visualization software created by Tableau Software.
- Tableau connects easily to nearly any data source
- Tableau allows for instantaneous insight by transforming data into interactive visualizations called dashboard.
- The company was founded in January 2003 by Christian Chabot, Pat Hanrahan and Chris Stolte, in Mountain View, California
- It is currently headquartered in Seattle, Washington, United States.
- Salesforce acquired Tableau on Aug 1, 2019.
- Every attribute categorized as dimension or measure in tableau.
- It offers drag and drop fields to create visualization.
- It provides powerful features to select and aggregate data like hierarchies, groups, sets, and filters.

Dashboards:

- It is powerful tool for mixing multiple views in a unique one.
- It provides interactive and common filter for all views grouped together.

Stories:

- It is a single sheet, which contains a sequence of worksheets or dashboards.
- It is powerful navigation tool for presentation

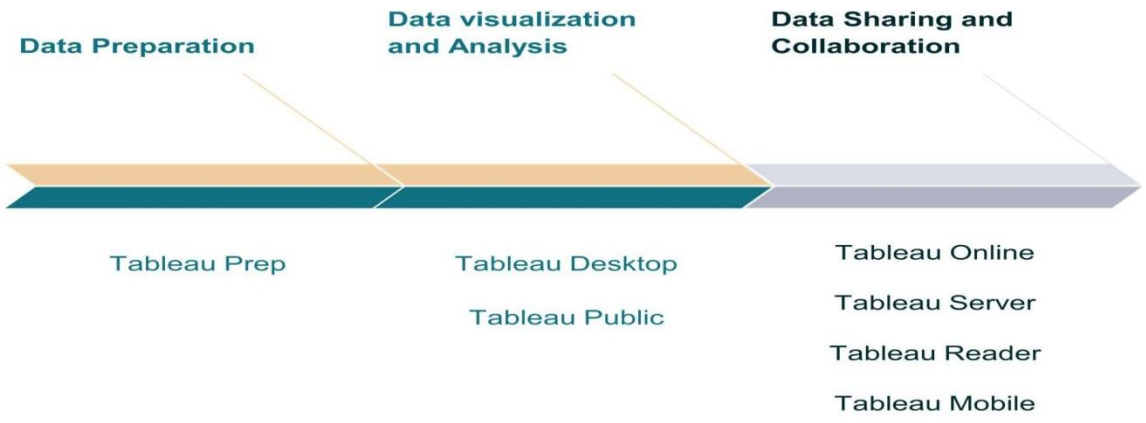
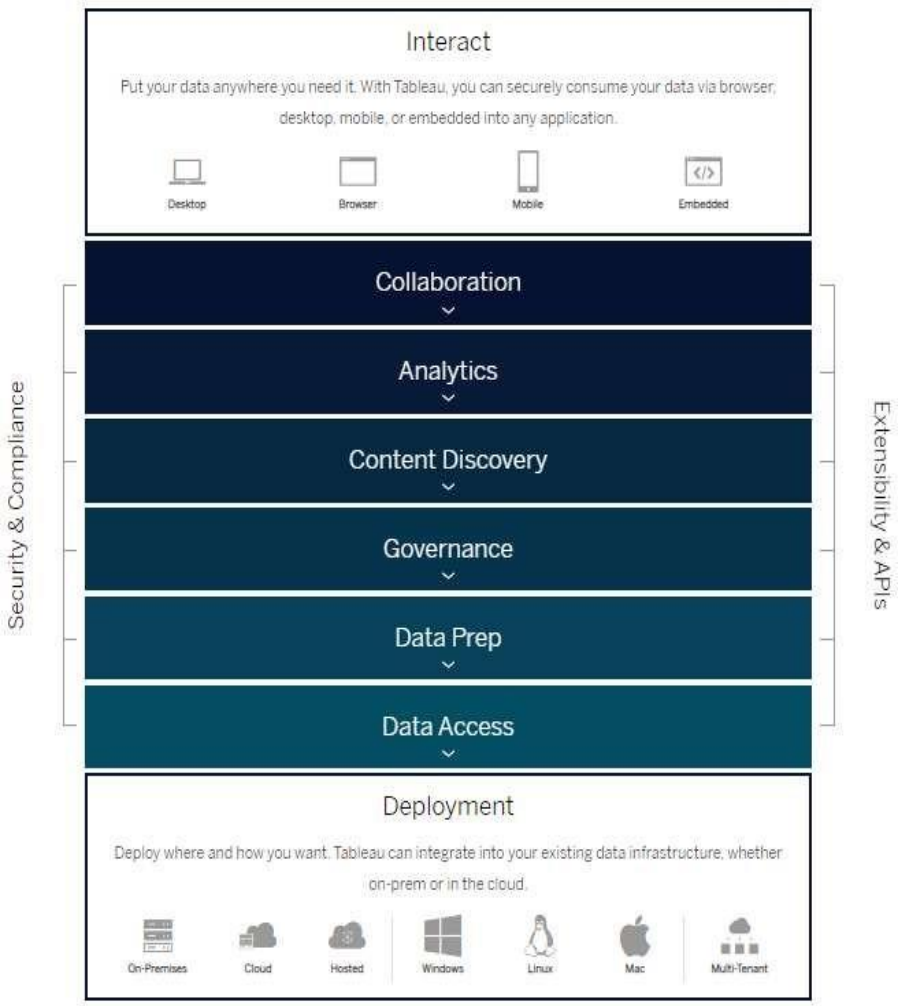


TABLEAU FEATURES



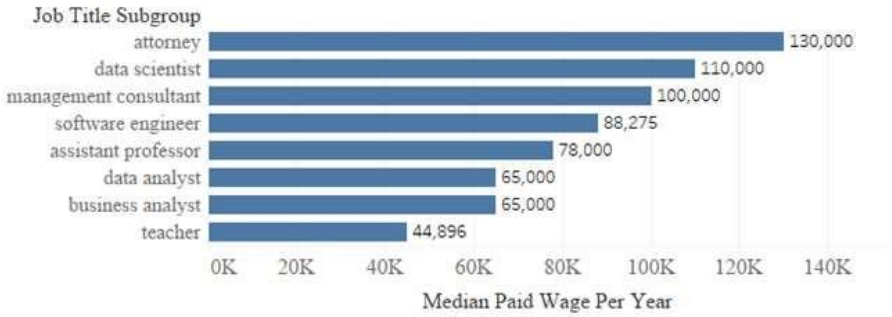
Variables considered for visualization

Dependent Variable: Paid Wage

Independent Variables: Case Received Date, Job title Subgroup, Employer, Work State

DATA VISUALIZATIONS

1. Median wage per year according to Job Title Subgroup



Summary:

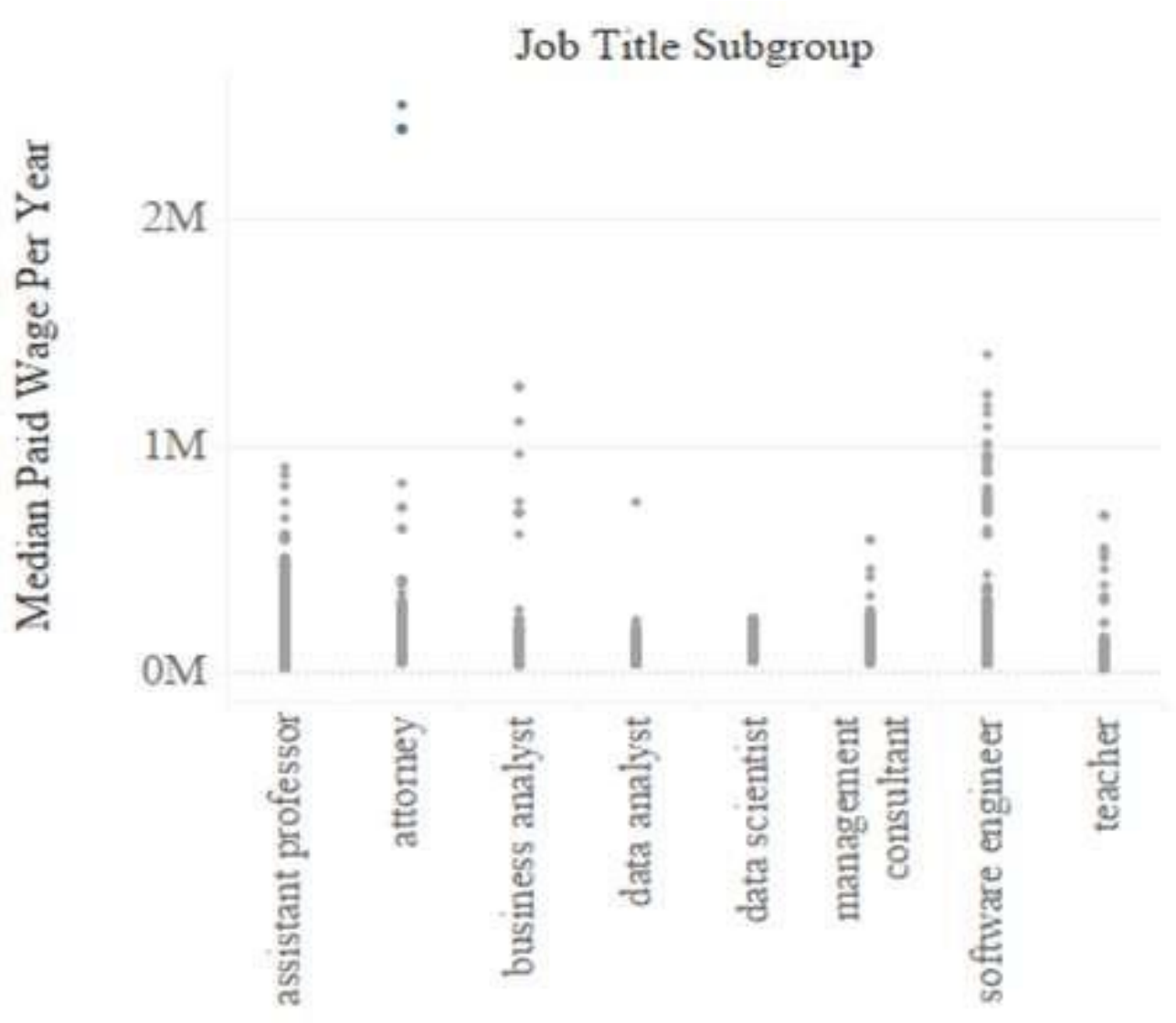
- We have shown the various median salaries of the selected job title subgroups with the help of bar charts.
- Attorney Job title subgroup is paid the highest median salary of 130,000 while teacher subgroup is paid the lowest salary out of all the other subgroups.
- We can conclude that **data scientist is the highest paid job title sub group** in case of data related jobs.

2. Median paid wage per year on the basis of Year in which case is received

Summary:

- This chart shows the median paid wage per year for each year from 2008 to 2015 as per the case received date with the help of bubble chart.
- The different color schemes are used for different job titles.
- Also, it shows that there are some outliers in case of attorney for year 2013 and 2014.

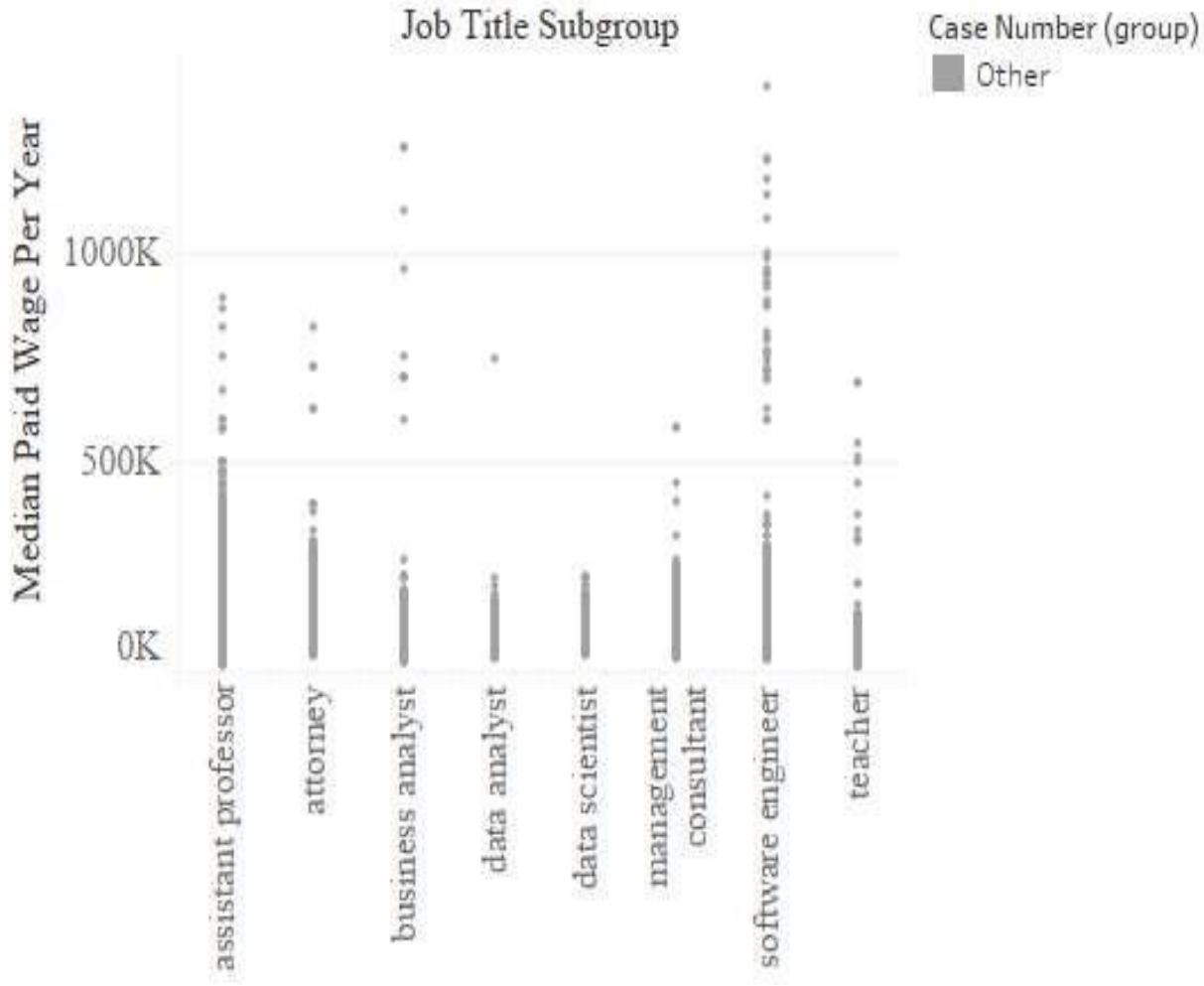
3. Median paid wage per Year per job title group showing Outliers



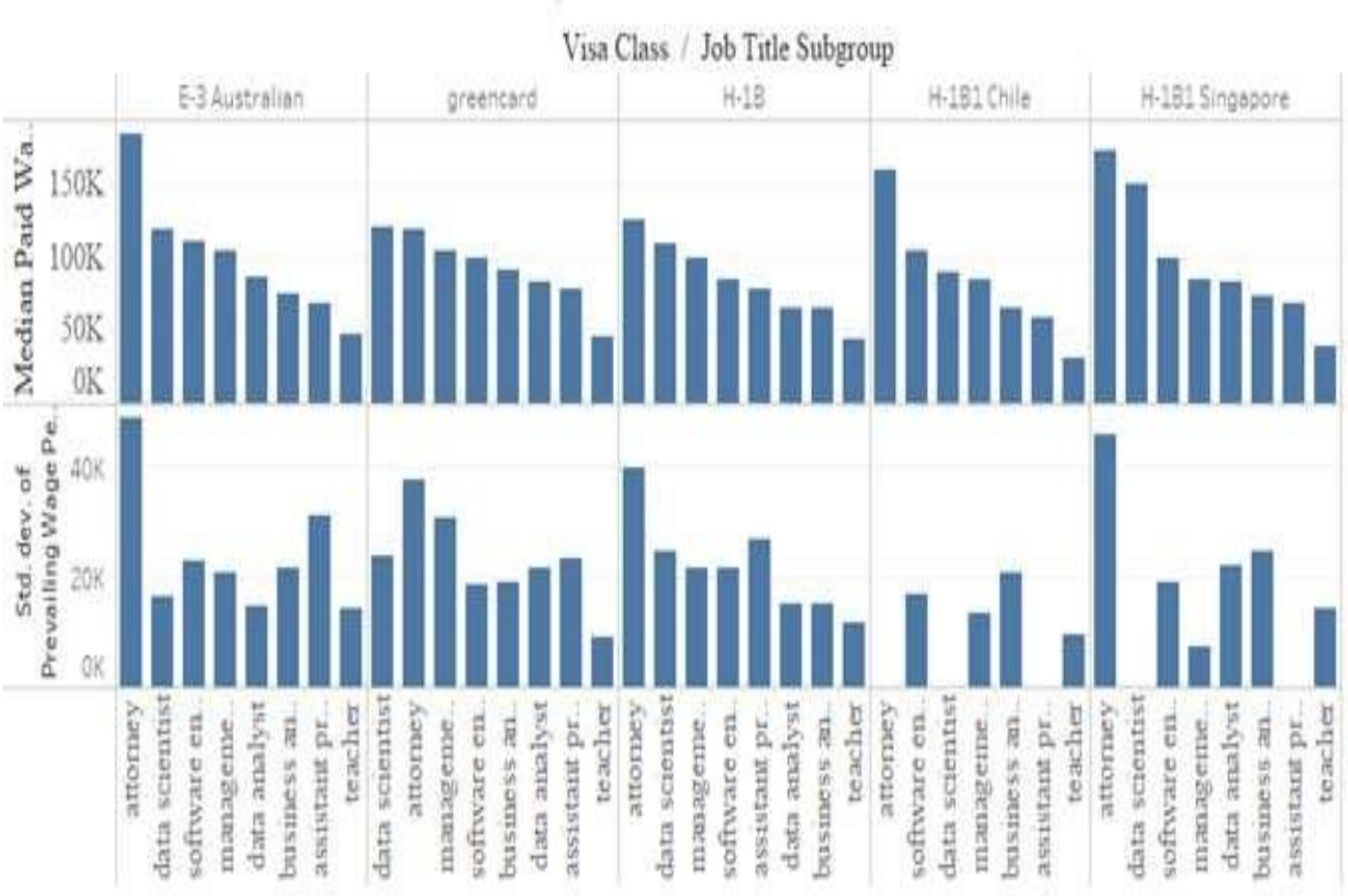
Summary:

- We have created the outlier group on the basis of case number to remove the outliers (shown in blue color) and then we selected the data other than the outlier group to create this outlier chart (using other).
- **Grouping in Tableau is grouping multiple members in a SINGLE dimension into a higher category and creating a set is grouping members from MULTI dimensions and/or condition into a dynamic or constant group.**

- In the second plot, we have filtered out the outlier group.



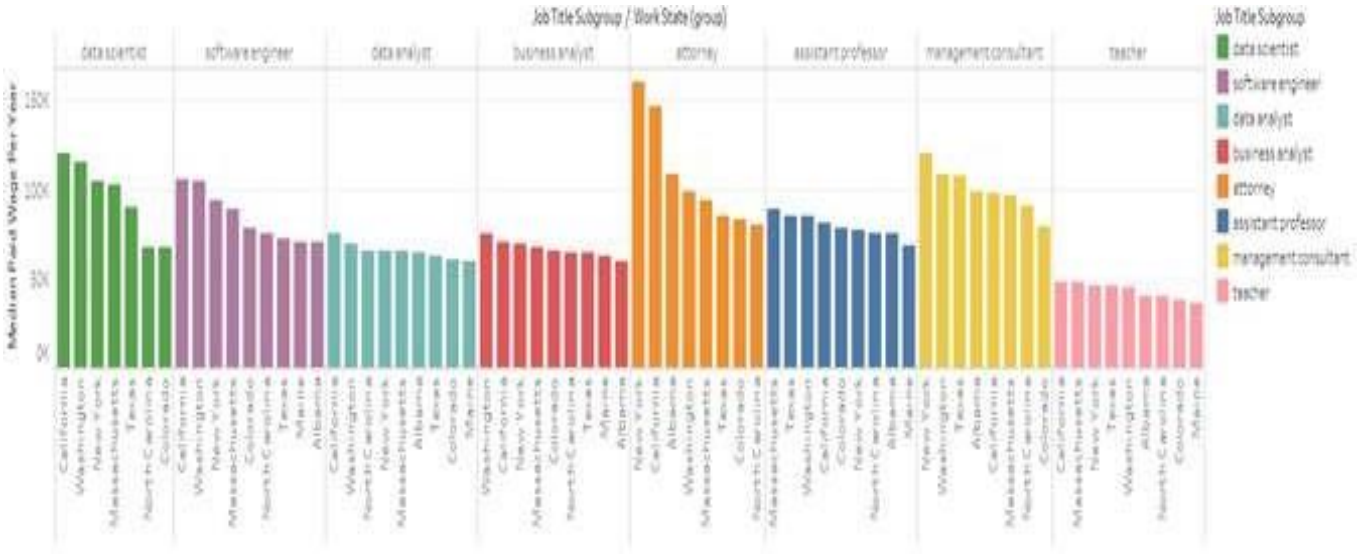
4. Influence of Visa Class on job title subgroups:



Summary:

- We are testing if the type of visa has any influence over the median paid wage of the various job title subgroups with the help median wage per year and Standard deviation of prevailing wage per year.
- The visa types we have taken are E-3 Australian Visa, Green card, H-1B, H-1B1 Chile and H-1B1 Singapore.
- We can see that there is not much difference caused by the visa type on the basis of median paid wage per year.
- Also, we have shown the standard deviation of prevailing wage of different job title subgroups in different visa types.
- This chart shows **visa type has no significant influence on paid wage.**

5. Median paid wage for different job title subgroups in different Work states



Summary:

- We are showing the median paid wage per year on the basis of job title subgroup which is filtered on the basis of state subgroup which consists of 10 states.
- We can see that California is showing the best median salary records across various job title subgroups.
- The colors used can be used for differentiating between the various job titles.

6. Median Paid wage per Year using Box plot

Summary:

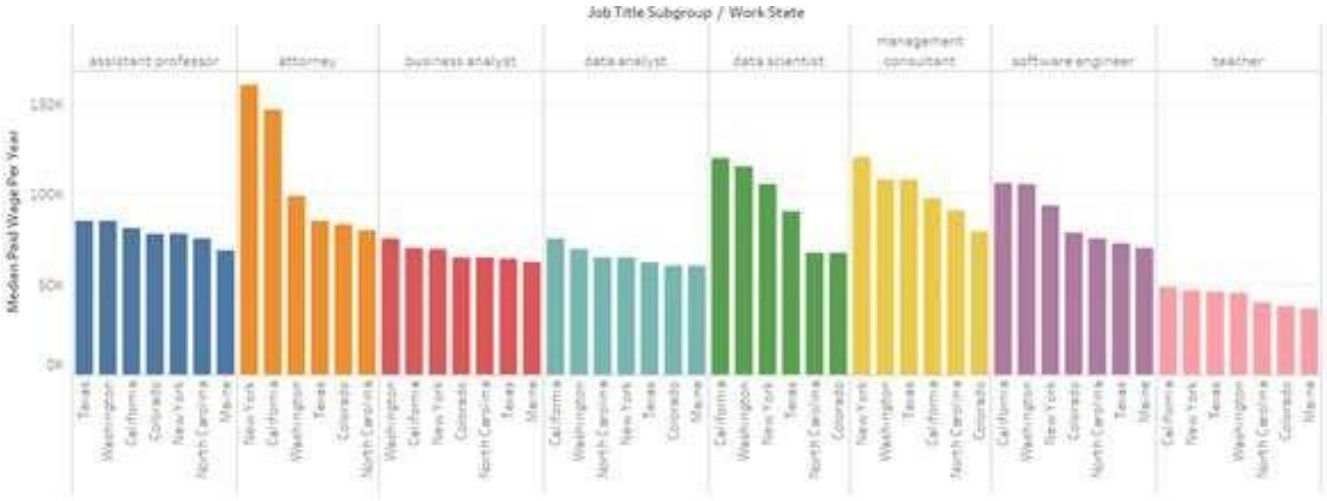
- A box plot displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum.
- In a box plot, we draw a box from the first quartile to the third quartile. A vertical line goes through the box at the median. The whiskers go from each quartile to the minimum or maximum.
- We can see that the median paid wage per year is decreasing with each passing year.

7. Median Paid wage per Year using Trend Line

Summary:

- These charts shows the median paid wage per year, maximum paid wage per year and minimum paid wage per year with respect to the case received year.
- Using trend line we will check if the line for a particular job title is significant or not.
- Trend line gives 3 data sets i.e equation, p value and r value in order to check the significance of a line.
- If our p value should less than 0.05 i.e 95 percent confidence interval in order for the line to be significant.
- Business analyst line is significant as its p value is less than 0.05.
- **Minimum paid wage is decreasing with time for data analyst, business analyst, and software engineer.**
- **Maximum paid wage is increasing with time for software engineer**

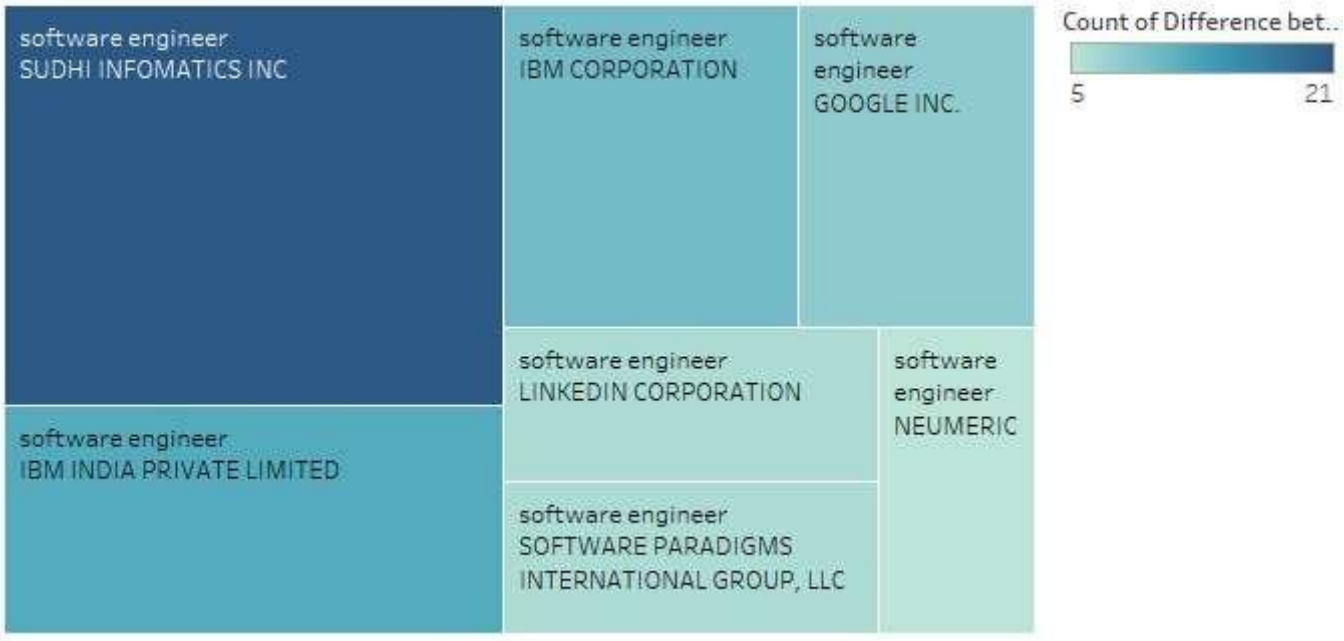
8. Paid wage per year for some selected Work states using calculation



Summary:

- We have used a calculation using if and else statement for using only the required state and dropping out the unwanted state.
- If the condition is satisfied then we plot the state and its data on the chart using view feature of tableau as a filter.
- If we select the no view option, the condition filter is ignored.
- Thus we can use this condition calculation to create the state group.
- Calculations allow you to create new data from data that already exists in your data source, as well as perform computations on your data. This allows you to perform complex analyzes and add fields to your data source on your own and on the fly.

9. Number of applicants being paid less than paid wage by the Employer for different job title subgroup



Summary:

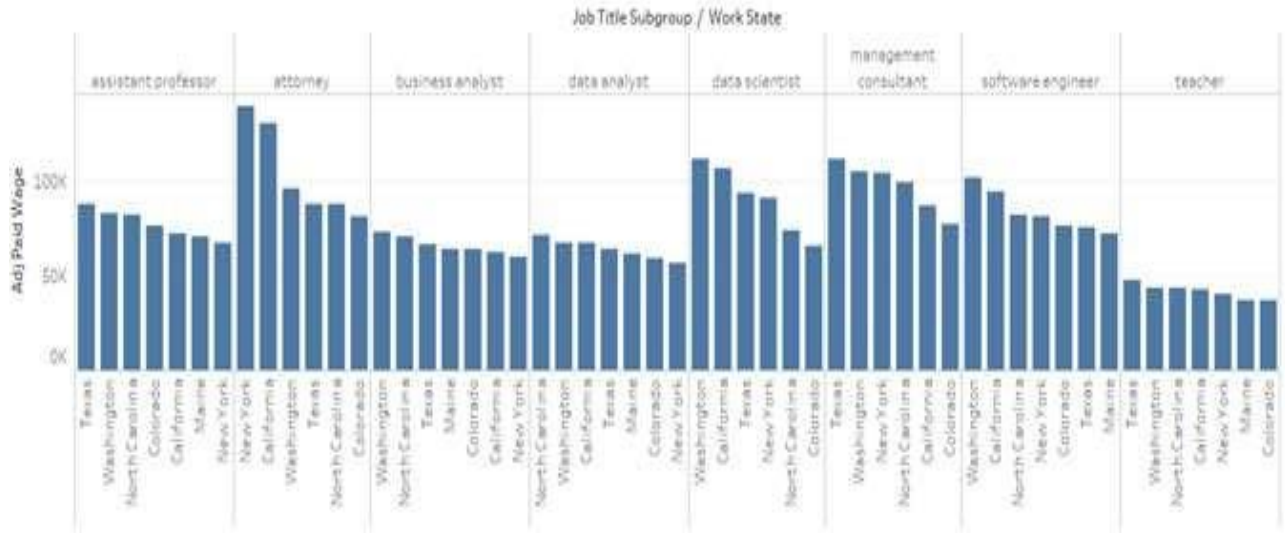
- We have used area chart to show the count of visa applicants for each employer in each job title subgroup having their prevailing wage less than paid wage.
- We have created this chart to see if the employers are paying less than the required wage.
- This chart is created using calculation that holds the median difference between the prevailing and paid wage for each job title subgroup.
- We have used the calculation “paid wage less than prevailing wage” as a filter for showing only zero and negative values.
- The above chart shows the **employers for software engineer job title subgroup**.

10. Adjusted Salaries on the basis of Cost of living

State						
Californ..	Colorado	Maine	New York	North Carolina	Texas	Washin..
94,162	64,189	60,528	70,970	61,930	62,379	87,547
112.30	102.20	97.70	115.30	91.70	96.70	103.20
89,938	70,450	69,076	70,247	76,336	67,218	97,268

We have checked that if the median wage per year is adjusted with the parity i.e is the cost of living. This data is filtered on the basis of chosen (interested) states.

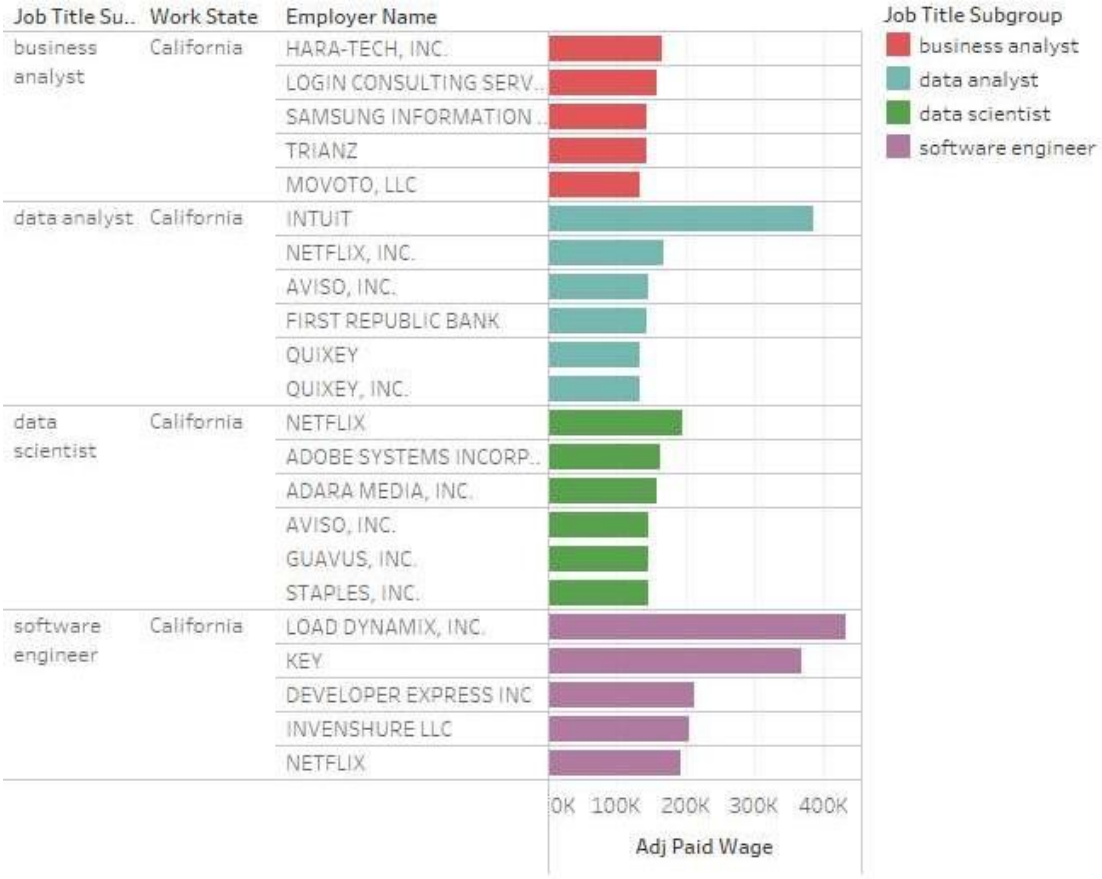
11. Highest paid Work States for each job title sub group using parameter



Summary:

- We have ranked the various states using a parameter TopX.(We can use the top or bottomrank categories)
- A parameter will allow you to provide a value to pass into Tableau. Parameters allow you to come up with scenarios or options that are not available in your data and create these values to put into your visualization. After creation, end users can control the input to see the results ofthe parameters effect.
- We have filtered the data using the parameter for the top 7 states paying adjusted paid wage.

12. Highest paying Employer in each state in each job title subgroup



Summary:

- We have ranked the various states using a parameter TopX.
- We have filtered the data using the parameter for the top 5 states paying adjusted paid wage.
- The color scheme shows the various job title subgroups.
- The above chart shows the top **5 employers in California for data related job title sub groups**.

CONCLUSION

Through this project we have learnt the application of Tableau to produce useful insights in the business world. It is secure, governed, scalable, and reliable, such that it has robust suite of built-in capabilities and it has the ability to integrate with systems. In this case, every attribute categorized as dimension or measure such that we can easily drag and drop fields to create visualization. Tableau offers powerful features to select and aggregate data like hierarchies, groups, sets, and filters. It consists of powerful tool for mixing multiple views in a unique one, dashboard. It provides interactive and common filter for all views grouped together.

We have also learnt the various steps of data preprocessing in this project. Our data set was collected from the official website of US department of labor. It was a large data set consisting of more than 16,000 rows of data. Our analysis was based on the median paid wage varying according to different job title sub group, year, work state and employer. The main focus was on four job title sub group i.e., data analyst, business analyst, data scientist, and software engineer and eight work states including California, Alabama, North Carolina, Colorado, Washington, Texas, New York, and Maine. With the help of above visualizations, one can easily choose the suitable job title sub group, work state and employer for data related jobs in US.