DS 0311-01 / Instructor: Norman Lo ( [lokman@mail.sfsu.edu](mailto:lokman@mail.sfsu.edu) ) / 07 October 2022

**DS 311 Group Assignment 1**

Tasks: Write a page of summary of the dataset includes 5 fun facts that your team found.  
The 1-page write up is not including your code or coding file.  
Push the coding file and summary report to the group project repository.

Fun facts about the dataset, option 2: “salary\_data\_states.xlsx”

After reading through the entire dataset once we realised that there were a lot of missing datapoints (“NaN”). We replaced those missing datapoints with “0”, “Unknown”, “Nothing” or “n”. *(Figure 5)*

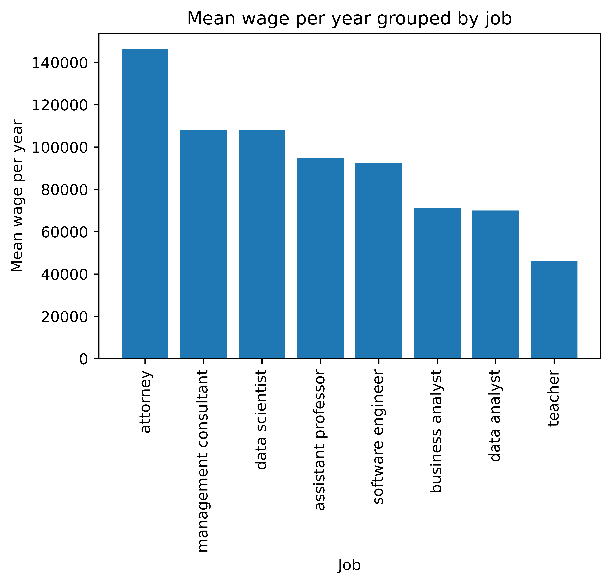
Fact number 1:

Figure Mean wage per year, grouped by job category

Sorting the mean wage per year by job subgroups, we can conclude that “attorney” related jobs make the most money, they make 140,000$ on average per year. The subgroup “teacher” is in last place and makes less than half of an attorney.

Fact number 2:

Ein Bild, das Text enthält.

Automatisch generierte BeschreibungSorting the mean wage per year by states, we can conclude that employees in “West Virginia” make the most money, they make around 109,000$ per year. We displayed the top 5 states in *(Figure 2)*

Figure Mean wage per year, grouped by state.

Fact number 3:

The mean salary per year in San Francisco is around 123,850$

Fact number 4:

Ein Bild, das Text enthält.

Automatisch generierte BeschreibungLooking at who makes the most money per year as a single person, we can see that there are a few attorneys that make substantially more than the rest in the dataset. Number 1 to 5 are attorneys that each make over a million more than number 6.

Figure Best paid people with job and city.

Fact number 5:

Ein Bild, das Text enthält.

Automatisch generierte BeschreibungConsidering the diversity within companies, we wanted to look at how salaries of those on work  
visas compare to green   
card holders. The data   
was sorted by job title  
subgroup, visa class, and lastly by organizing   
the mean paid wage per year. Across the board we note that those not holding green cards are paid significantly less than green card holders. The most notable jump is a data scientist from Chile gets paid a mean annual wage of 90000.00   
compared to a greencard holder who makes 116507.76, the amount of experience   
would have to also be analyzed to gather more conclusive results. When isolating   
wage by visa status we can notably see that greencard holders on average make   
more than those on work sponsored visas.

Figure Visa class vs paid wage

Summary:

The data we have shown is of course only representative of the given dataset and should not be understood in general terms.

To summarize our first look at the dataset, we can say that the data provided contains quite a bit of intel about deployment data, but is also very broad and provides a lot of information about its relationship/relevance to visa status, the state and city where one works, etc....

**Code screenshots attachment:**

Ein Bild, das Text enthält.

Automatisch generierte Beschreibung

Figure replacing missing data

Ein Bild, das Text enthält.

Automatisch generierte Beschreibung

Figure Code for fact 1

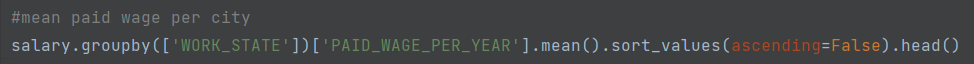


Figure Code for fact 2



Figure Code for fact 3

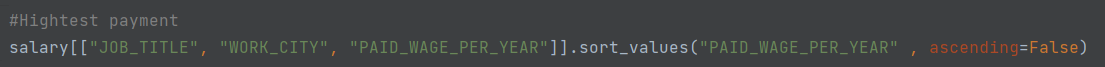


Figure Code for fact 4

Ein Bild, das Text enthält.

Automatisch generierte Beschreibung

Figure Code for fact 5