Big Data, Big Money

Abstract

Welcome to Rihans first lab. I am going to be looking at data on Hourly Wage in Seattle.

We are looking at the wage per hour in Seattle City Utilities. We found the data from Seattles data website.

Some questions I'm looking to answer:

- What job pays the most money in Seattle Utilities?
- What job pays the least money in Seattle Utilities?
- What is the youngest age of someone who works in Seattle Utilities?
- What is the oldest age of someone who works in Seattle Utilities?

Dataset Preparation

The dataset is from Seattle City Gov (https://data.seattle.gov/City-Business/City-of-Seattle-Staff-Demographics/5avq-r9hj (https://data.seattle.gov/City-Business/City-of-Seattle-Staff-Demographics/5avq-r9hj)). The dataset shows all current City of Seattle Employees. It was made in February and was last updated on August 30th, 2019. It is updated monthly and includes data on about 14,200 employees. The data set has seven columns that include race, sex, department, age, hourly rate ect. I will look at hourly wage and age to see if older people are paid more.

This is code that opens the file and calls it demo data.

In [40]:

```
#open the file
data_file = open("City_of_Seattle_Staff_Demographics.csv", "r");

#create an empty list to store the data from the csv file
demo_data = [];

#put all lines from the file into a list
for row in data_file:
    demo_data.append(row);

#close the file. It's the courteous and clean thing to do
data_file.close();

# show the first five items in the list. Notice that they are all strings
# the \n part of the string means "newline" - its the symbol that represents the en
print(demo_data[:5]);
```

```
['Race/Ethnicity,Sex,Department,Age,Hourly Rate,Regular/Temporary,Empl
oyee Status\n', 'Hispanic or Latino,M,Municipal Court,57,29.75,R,A\n',
'Hispanic or Latino,F,Municipal Court,39,27.18,R,A\n', 'Hispanic or La
tino,M,Police Department,52,59.75,R,A\n', 'Hispanic or Latino,M,Fire D
epartment,30,38.28,R,A\n']
```

Now I am going to pull out two columns, age and wage. Age is an integer telling us how old someone is and wage is the dollar amount someone makes. I am interested in these two columns because I am looking at how age may or may not affect wage.

In [44]:

```
#remove the first element (column header)
# the header is useful for telling us what the columns mean,
# but you don't want to include it in your mathematical calculations
del demo_data[0];

# create two empty lists - one for age and one for wage demo data
# we want to make a list of just the age data and just the wage data
age = [];
wage = [];

# loop (go through every row one by one) through the original demo_list
for item in range(len(demo_data)):

    # this line is a bit confusing, but what it is doing is splitting out each comm
    # from a row and turning it into a list
    # for example, this line of code turns he single string '57, 29.75\n'
    # into two separate strings '57' and '29.75'
    # After being split, the two separate strings are then stored in another list c
    demo_row = demo_data[item].split(",")

    #now that we've split the row into its two parts, we can pick and choose
```

```
    # add the second to specific lists for age and wage
    age.append(int(demo_row[3]))
    wage.append(float(demo_row[4]))

agemin = min(age)

agemax = max(age)

wagemax = max(wage)

wagemin = min(wage)




print('minimum age:')
print(agemin)
print("maximum age:")
print(agemax)
print("maximum wage:")
print (wagemax)
print ("minimum wage:")
print (wagemin)
```

```
minimum age:
14
maximum age:
92
maximum wage:
162.8353
minimum wage:
5.53
```

Minimum age is 14.

Maximum age is 92.

Maximum wage is 162.84

Minimum wage is 5.53

Now I am going to find the average of age and wage from the data set.

Data Modeling

In [42]:

```
# to count how many items are in the list, use the function len

someNbrs = age
count = len(someNbrs)
print("Nbr of age is ", count)

# to add up all the items in the list, use the function sum
total = sum(someNbrs)
print("Total sum of age is ", total)

# you can then use these two values to calculate the average (also called the mean)
avg=total/count
print("Average of age is ", avg)

# or, use the function statistics.mean to do the calculations
# you must include the import statistics line first
import statistics
another_mean = statistics.mean(someNbrs)
print("Mean (average) of values age is", another_mean)
```

```
Nbr of age is  14194
Total sum of age is  632115
Average of age is  44.53395801042694
Mean (average) of values age is 44.53395801042694
```

In [43]:

```
# to count how many items are in the list, use the function len

someNbrs = wage
count = len(someNbrs)
print("Nbr of wage is ", count)

# to add up all the items in the list, use the function sum
total = sum(someNbrs)
print("Total sum of wage is ", total)

# you can then use these two values to calculate the average (also called the mean)
avg=total/count
print("Average of wage is ", avg)

# or, use the function statistics.mean to do the calculations
# you must include the import statistics line first
import statistics
another_mean = statistics.mean(someNbrs)
print("Mean (average) of wage is", another_mean)
```

```
Nbr of wage is  14194
Total sum of wage is  568240.0329000067
Average of wage is  40.03381942370063
Mean (average) of wage is 40.03381942370015
```

Data Analysis and Conclusions

I found that the oldest person working in Seattle City Utitilies is 92 years old, I am wondering what they could possibly be doing at that age and how much money they are making. The youngest person is 14 years old, and I am wondering if they are in school while working. The minimum wage is 5.53 dollars, I am wondering what job pays so little. The maximum wage is 162.84 dollars, and I am wondering how old the person making this much money is and what job they currently have.

Further Exploration:

- Is there a wage gap between men and women?
- Is there a wage gap between people of color and white people?
- What job(s) pay the most?
- Is the pay suitable for the job?
- How many people are making more than 20 dollars?
- What job is the most dangerous and how much do those people make?

Awknowledgements

I have to awknowledgement! First, Ziah helped me figure out how to put the maximum and minimum age, once

she helped me with that I was able to figure out the minimum and maximum wage. I was also able to figure out the average of both wage and age. Ms. Scoyners gave me her code to use and helped me set up my first notebook! Because of her help I was actually able to do the assignment.

In [ ]: