Analysis of Citywide Payroll Data

Team12: Ryan Austin Yon, Yunhan Zhang

IST652: Scripting for Data Analysis

Pro. Carlos Caicedo Bastidas

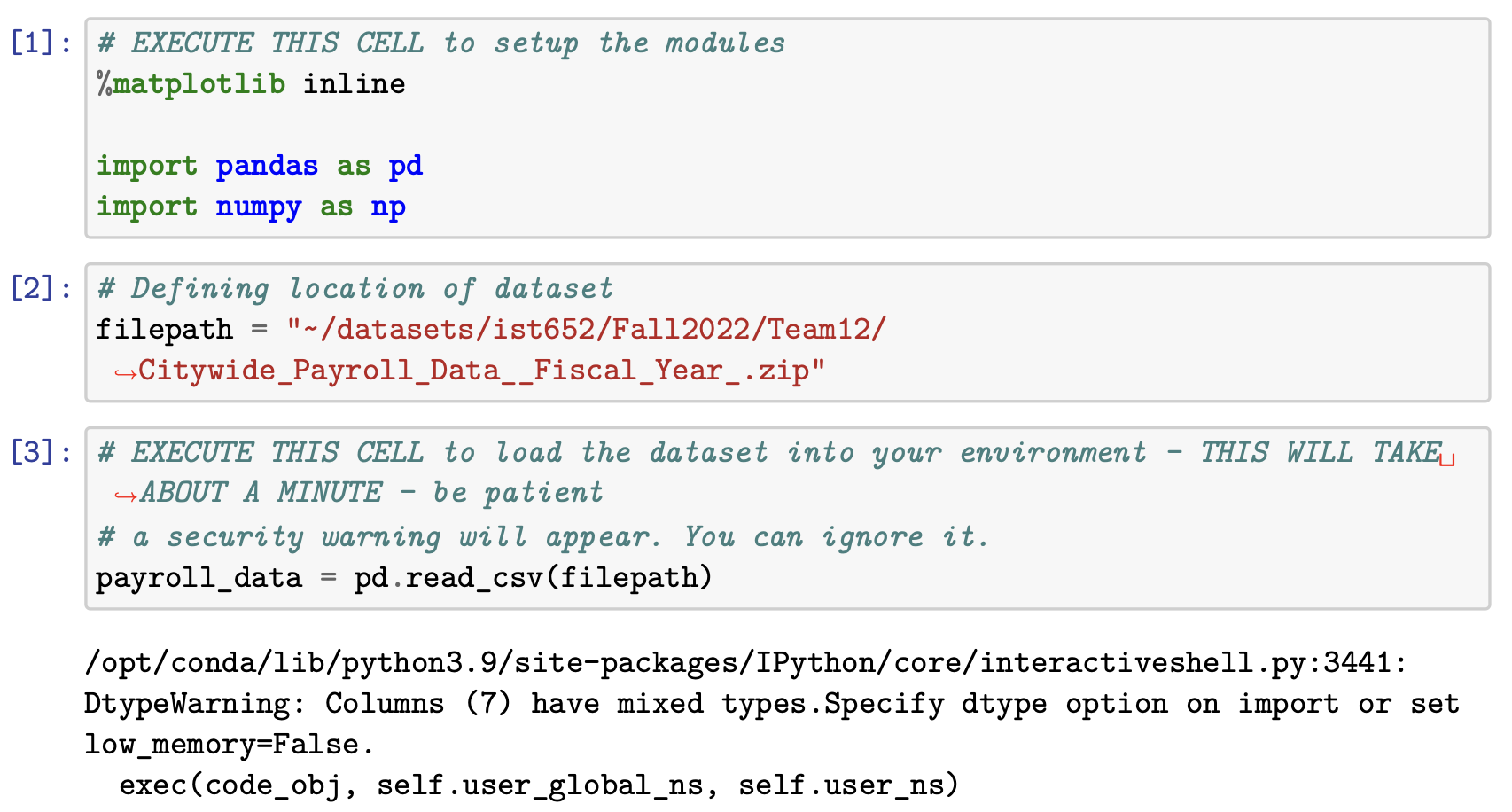
Dec 7, 2022

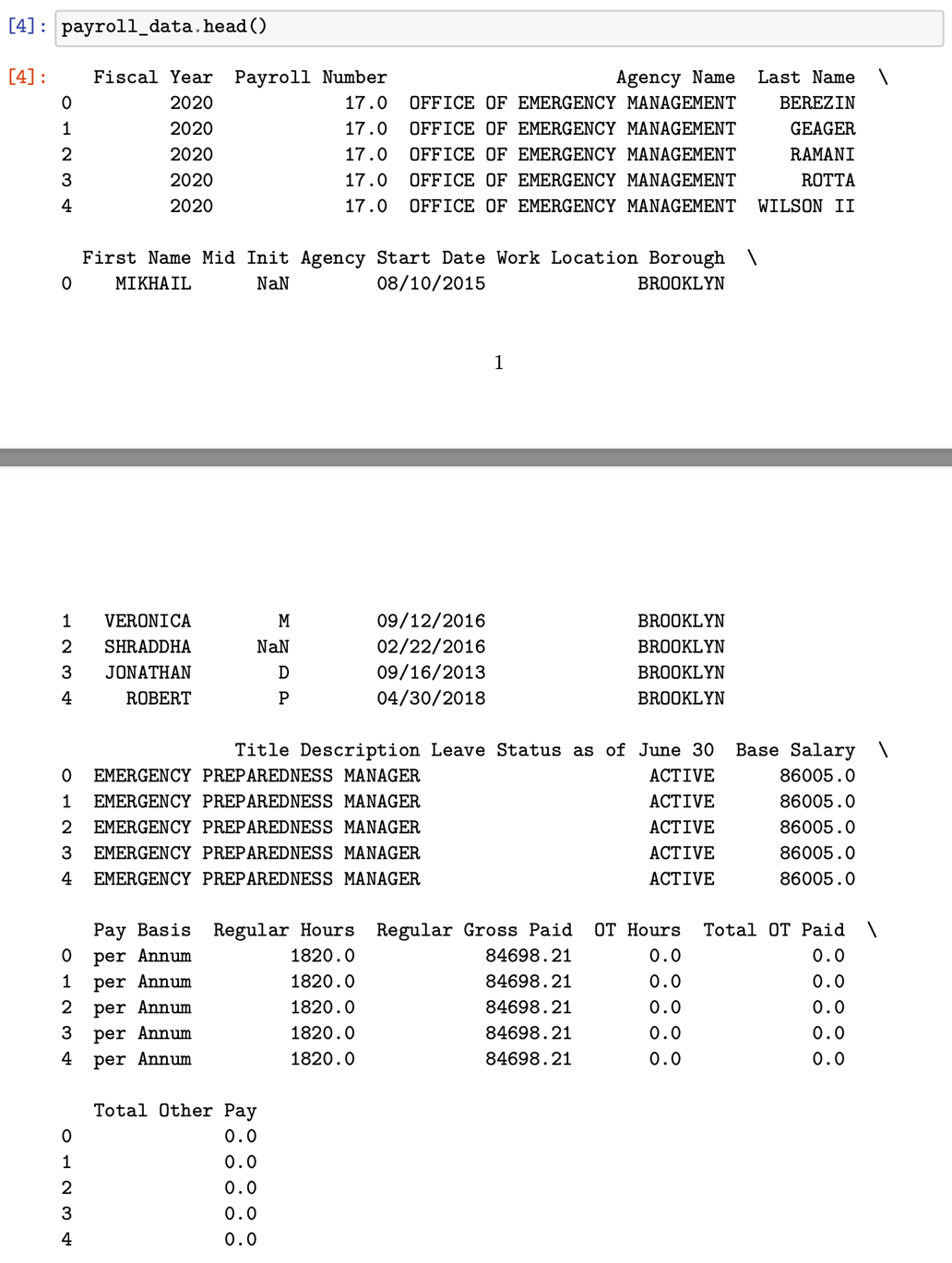
**1. Data Analysis Objectives**

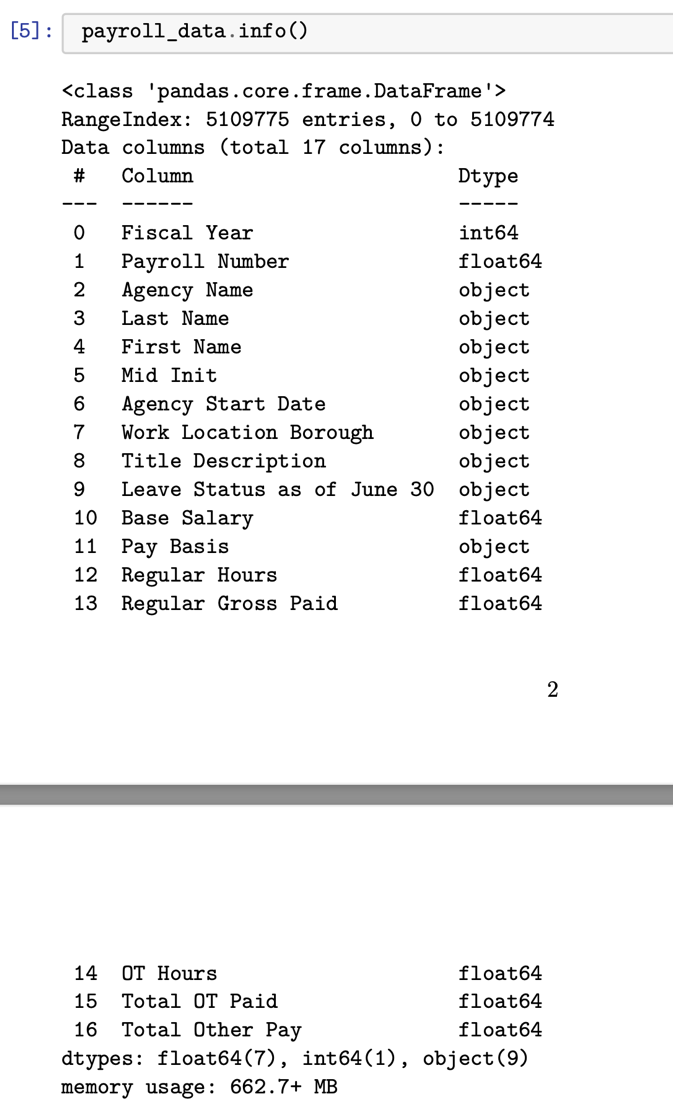
Our goal is to analyze the characteristics of each element and the connection between each other through the existing information from Citywide Payroll Data (Fiscal Year), to provide constructive analysis before and post COVID-19 pandemic. This analysis can be handed to local government departments or synchronized with the public to let them better understand the economics during the period of time.

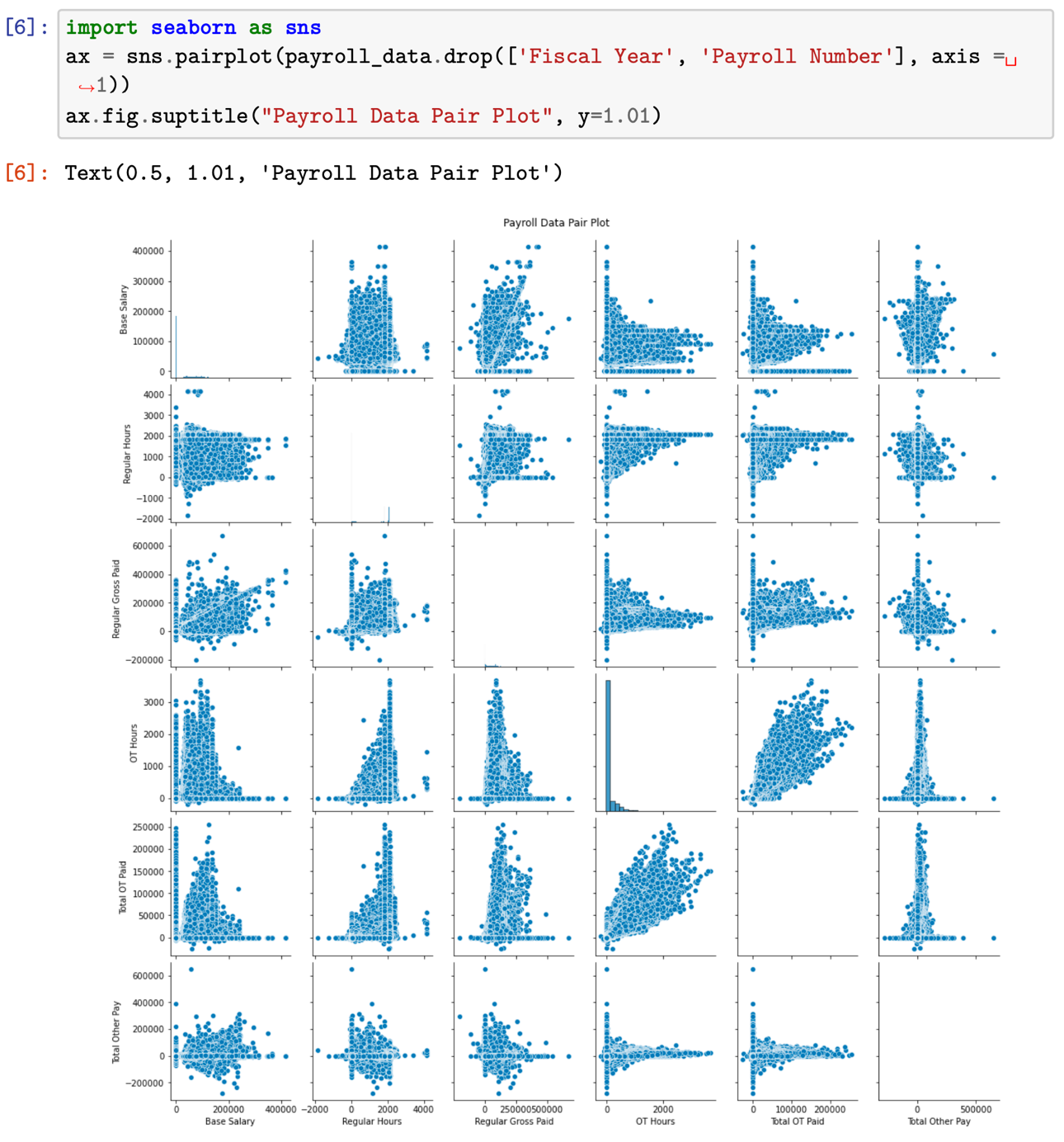
**2. Data Set**

Data is collected because of public interest in how the City’s budget is being spent on salary and overtime pay for all municipal employees. Each record represents the following statistics for every city employee: Fiscal Year (2014 – 2022), Agency, Last Name, First Name, Middle Initial, Agency Start Date, Work Location Borough, Job Title Description, Leave Status as of the close of the FY (June 30th), Base Salary, Pay Basis, Regular Hours Paid, Regular Gross Paid, Overtime Hours worked, Total Overtime Paid, and Total Other Compensation. This data can be used to analyze how the City's financial resources are allocated and how much of the City's budget is being devoted to overtime.







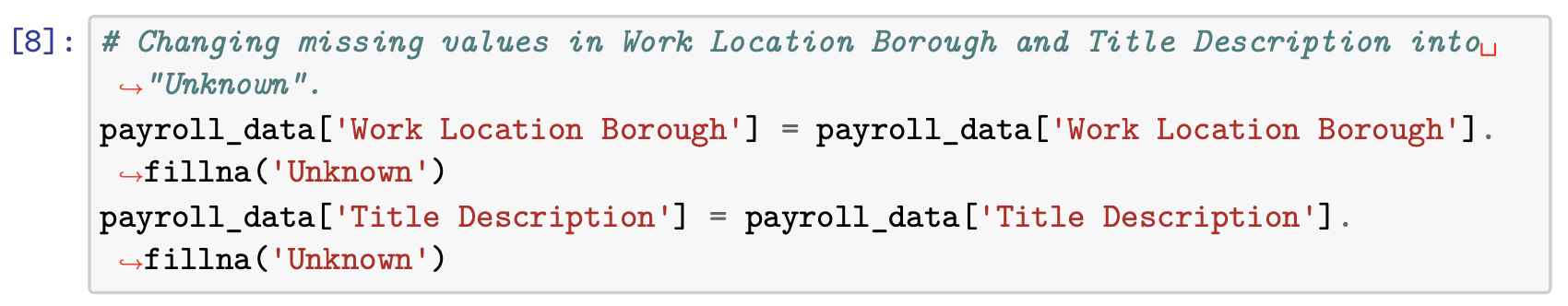


**3. Data Dictionary**

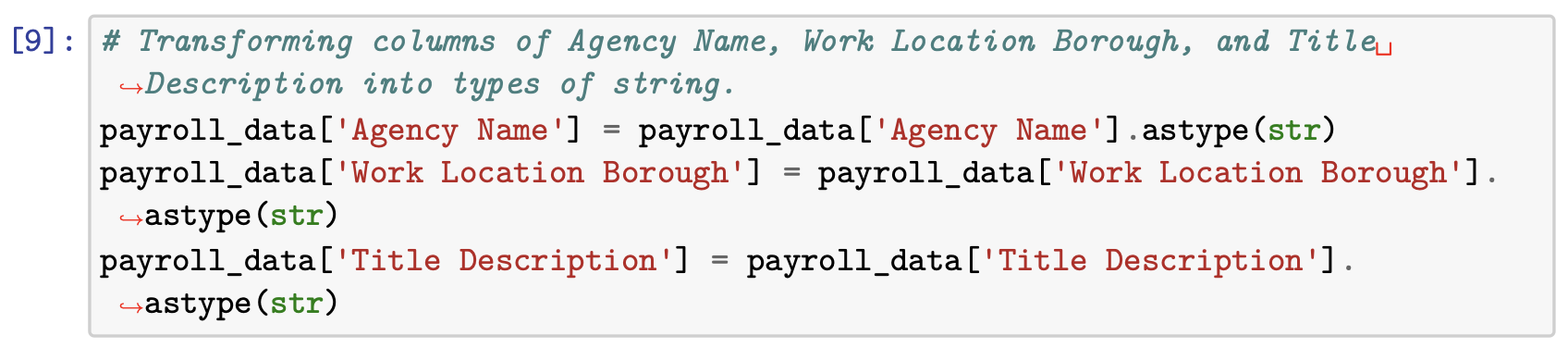
* Fiscal Year: Fiscal Year (2014 – 2022);
* Payroll Number: Payroll Number;
* Agency Name: The Payroll agency that the employee works for;
* Last Name: Last name of employee;
* First Name: First name of employee;
* Mid Init: Middle initial of employee;
* Agency Start Date: Date which employee began working for their current agency;
* Work Location Borough: Borough of employee’s primary work location;
* Title Description: Civil service title description of the employee;
* Leave Status as of June 30: Status of employee as of the close of the relevant fiscal year: Active, Ceased, or On Leave;
* Base Salary: Base Salary assigned to the employee;
* Pay Basis: Lists whether the employee is paid on an hourly, per diem or annual basis;
* Regular Hours: Number of regular hours employee worked in the fiscal year;
* Regular Gross Paid: The amount paid to the employee for base salary during the fiscal year;
* OT Hours: Overtime Hours worked by employee in the fiscal year;
* Total OT Paid: Total overtime pay paid to the employee in the fiscal year;
* Total Other Pay: Includes any compensation in addition to gross salary and overtime pay.

**4. Data Cleaning**

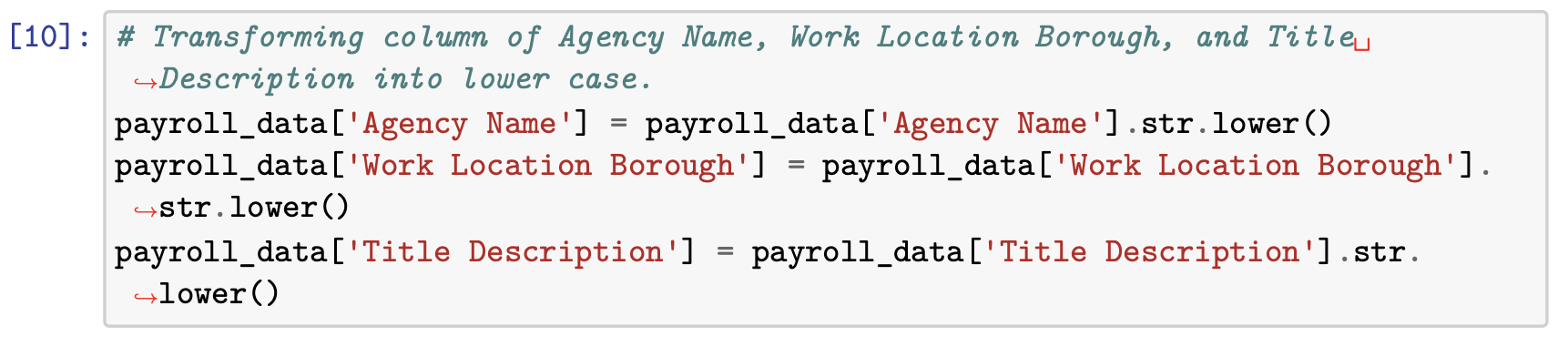
For the missing values, we noticed that there are many missing values across each column in the dataset; however, we will just focus on the missing values in ‘Work Location Borough’ and ‘Title Description’. Therefore, we changed these missing values into ‘Unknown’.



We also transformed columns of ‘Agency Name’, ‘Work Location Borough’ and ‘Title Description’ into String Types.



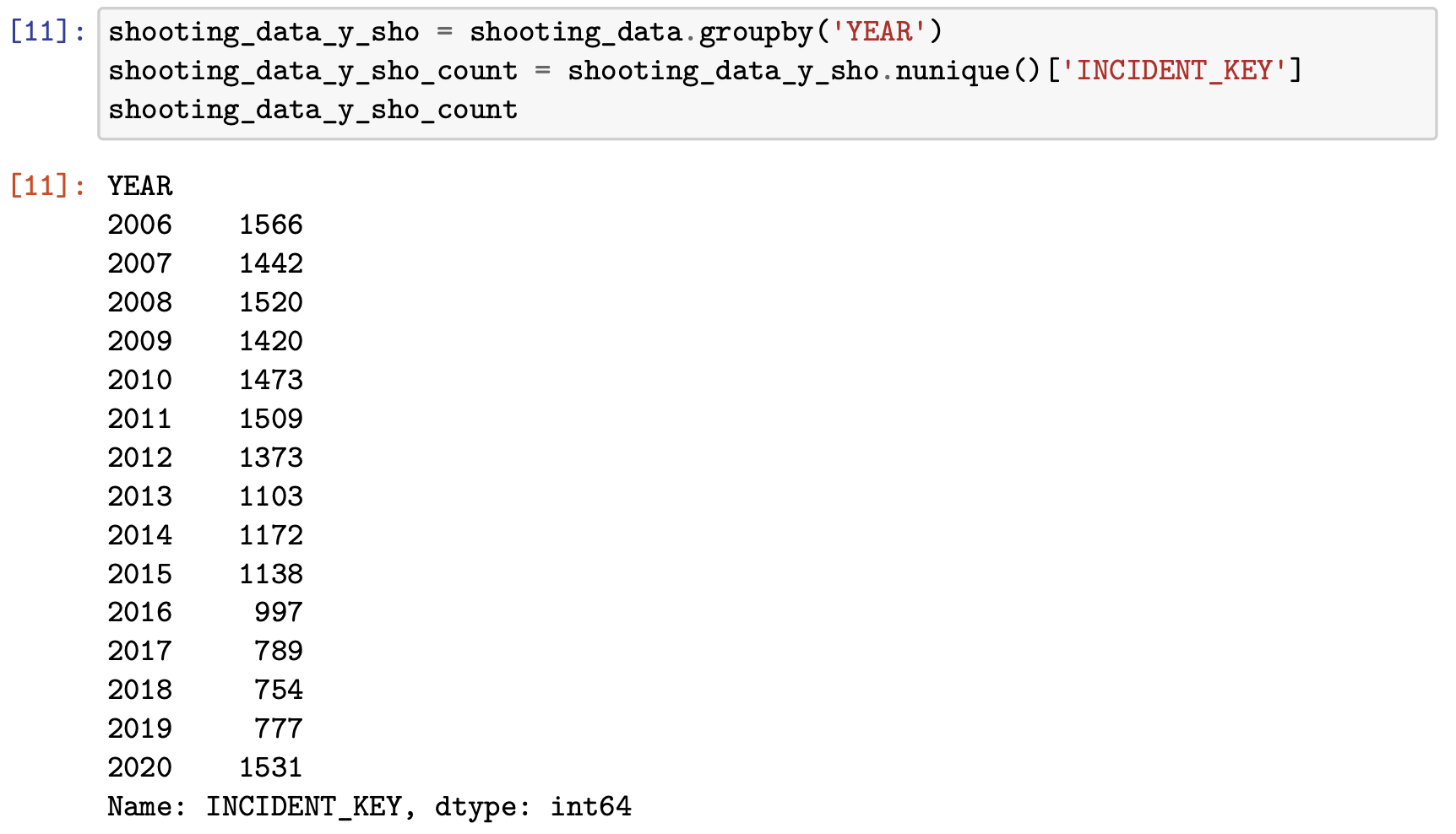
To observe these columns clearly, we transformed these into lower case.



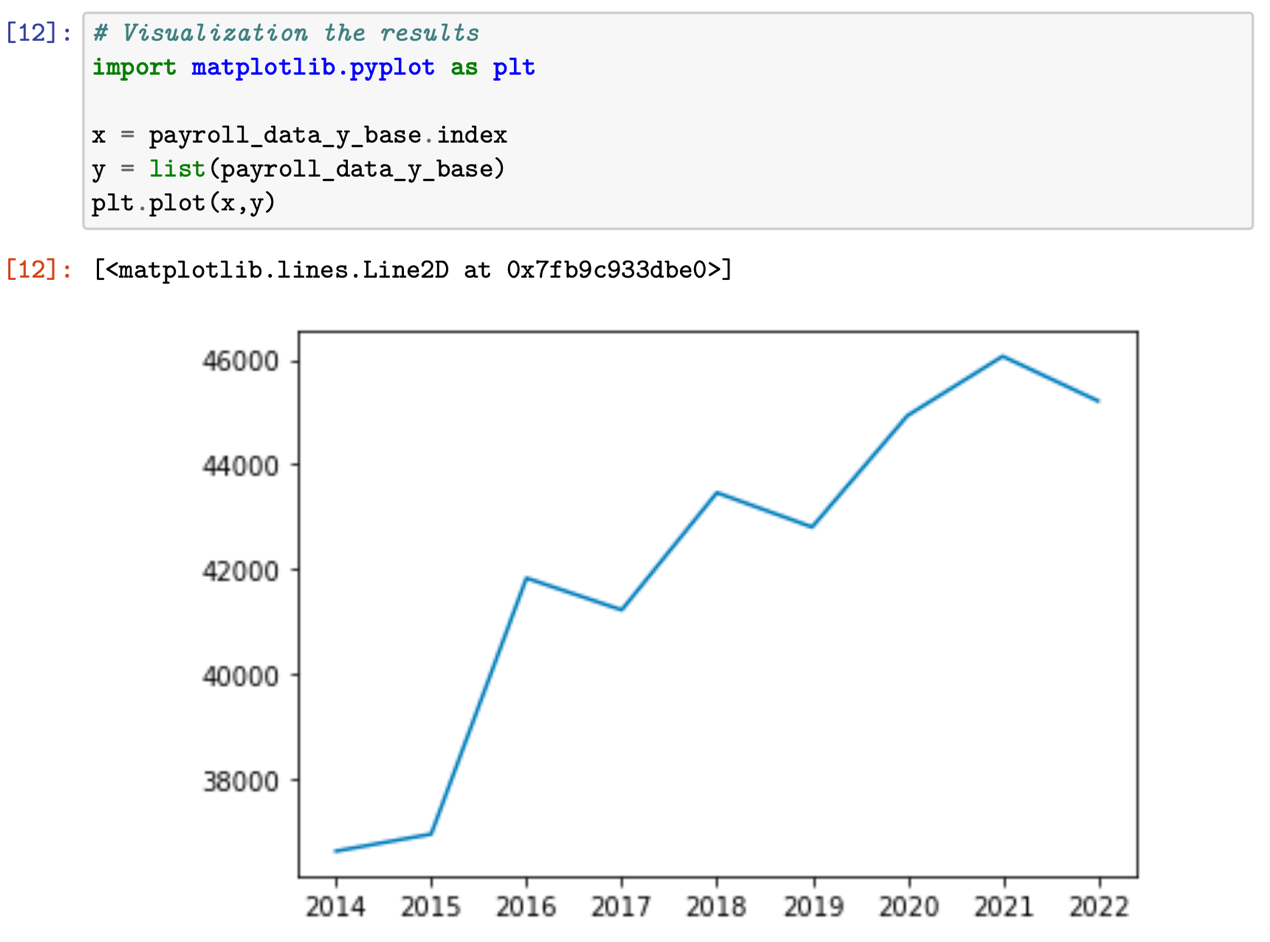
**5. Data Analytics**

**5.1 Trend change in median base salary over years**

a. we used the following codes to calculate the median of base salary in each year from 2014 to 2022:



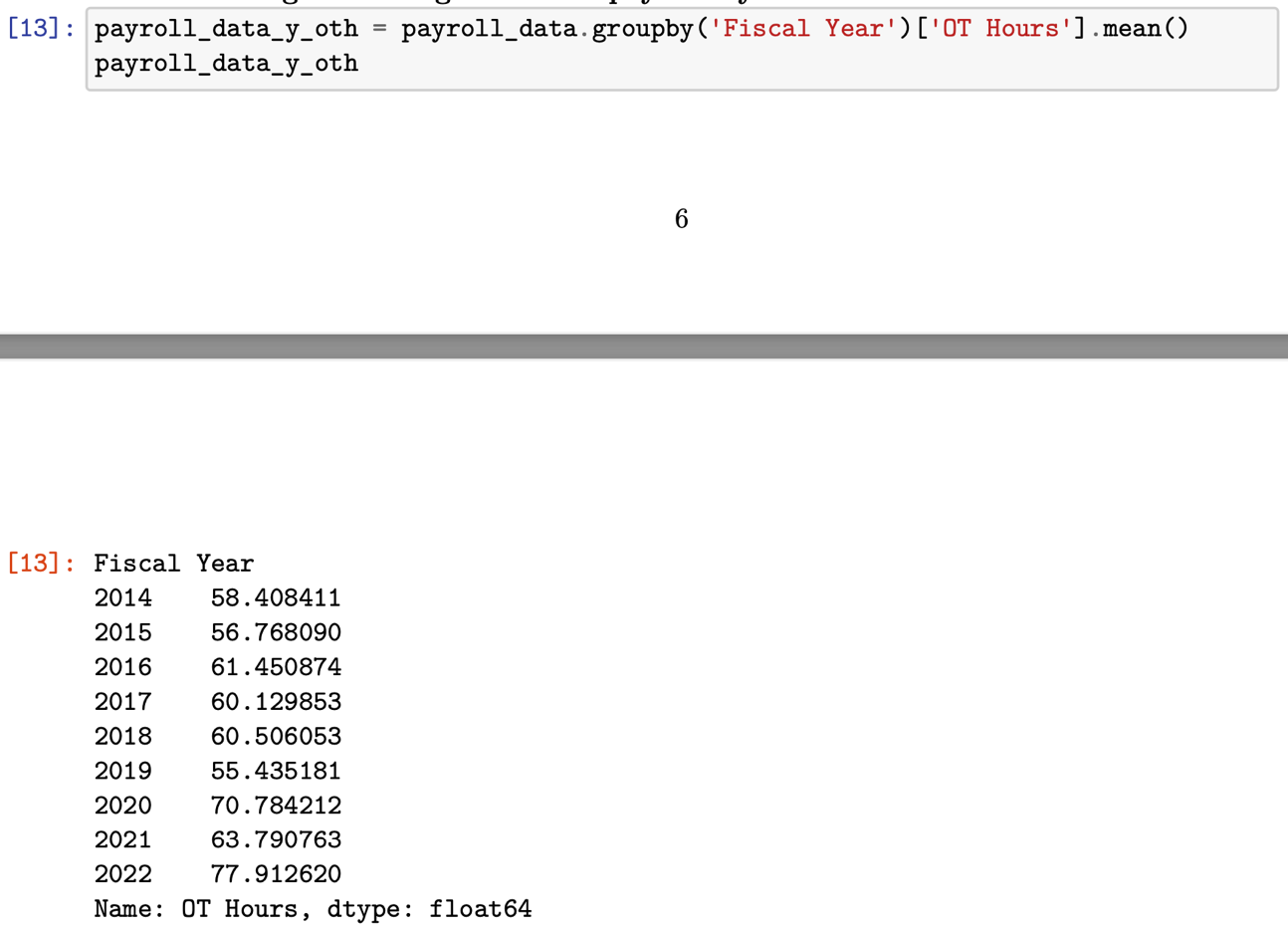
b. And, the following are the visualization of the median of base salary in each year:

****

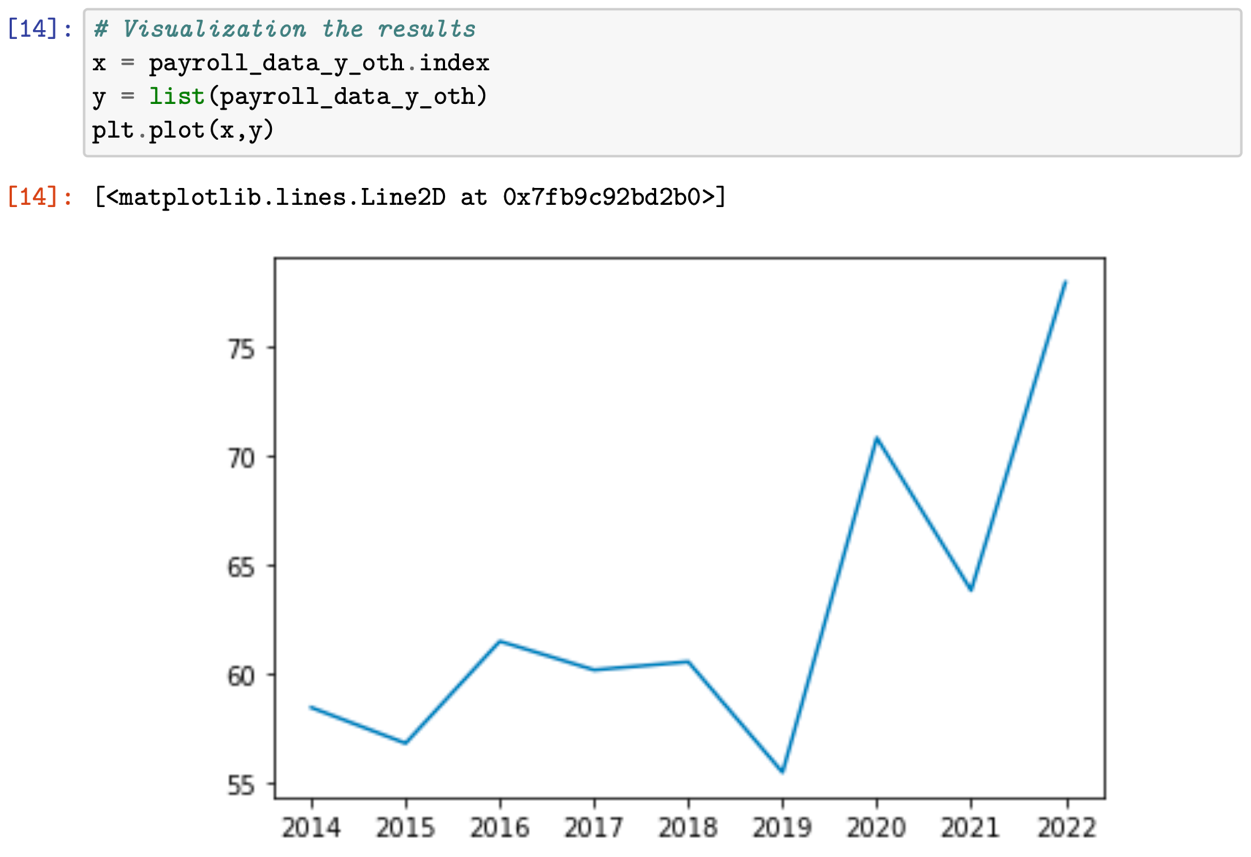
This plot shows that the base salary overall goes up since 2014. However, as we pay attention to recent years (2021-2022), it has a trend of going down. After we reviewed some news and reports in the past few years, we found that the COVID-19 pandemic did not affect people’s base salary that much. This might be related to the relevant decrees issued by the government during the COVID-19 pandemic.

**5.2 Trend change in average overtime hours over years**

a. we used the following codes to calculate the average overtime hours in each year from 2014 to 2022:

****

b. And, the following are the visualization of the average overtime hours in each year:



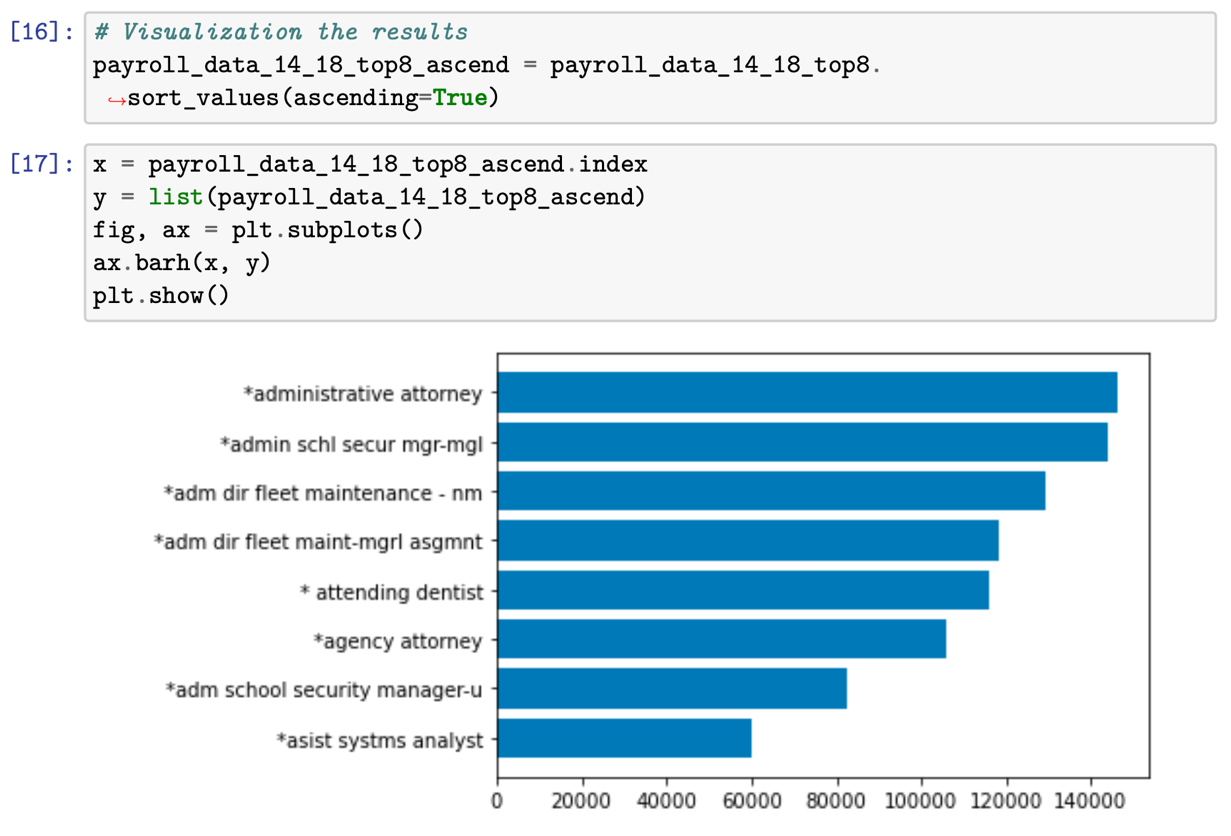
This plot shows that the total overtime hours was stable before COVID-19 pandemic (2014- 2018). And the total overtime surge to a high level after COVID-19 pandemic (2019-2020). It went down a little from 2020 to 2021 and went up to high again from 2021 to 2022. Our guess is the agencies in this dataset may require over time during the COVID-19 pandemic.

**5.3 Job Titles that have the highest amount of base salary from 2014 to 2018**

a. we used the following codes to calculate the highest amount of base salary from 2014 to 2018:

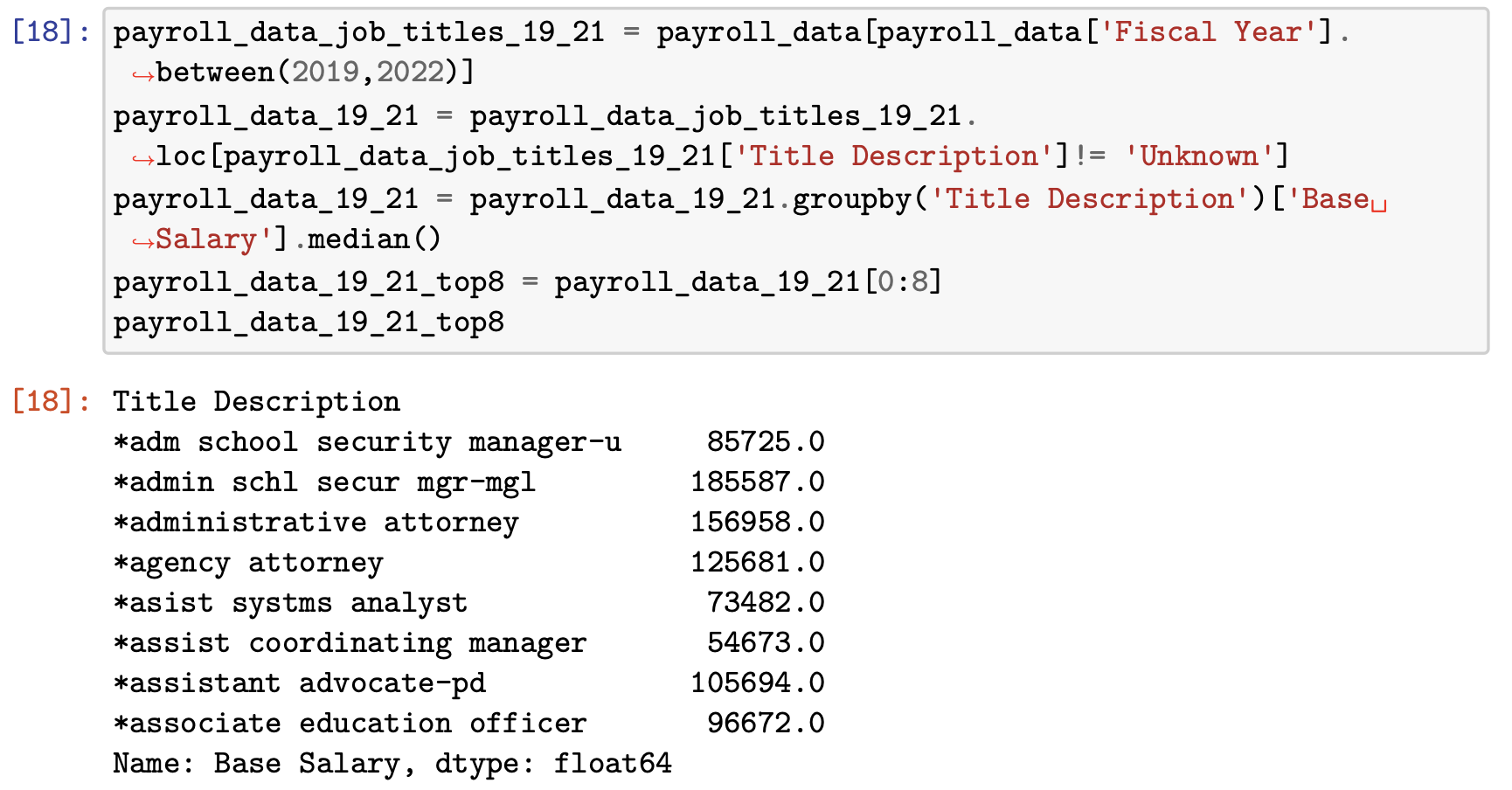


b. And, the following are the visualization of the highest amount of base salary from 2014 to 2018 (before COVID-19):



**5.4 Job Titles that have the highest amount of base salary from 2019 to 2021**

a. we used the following codes to calculate the highest amount of base salary from 2019 to 2021:



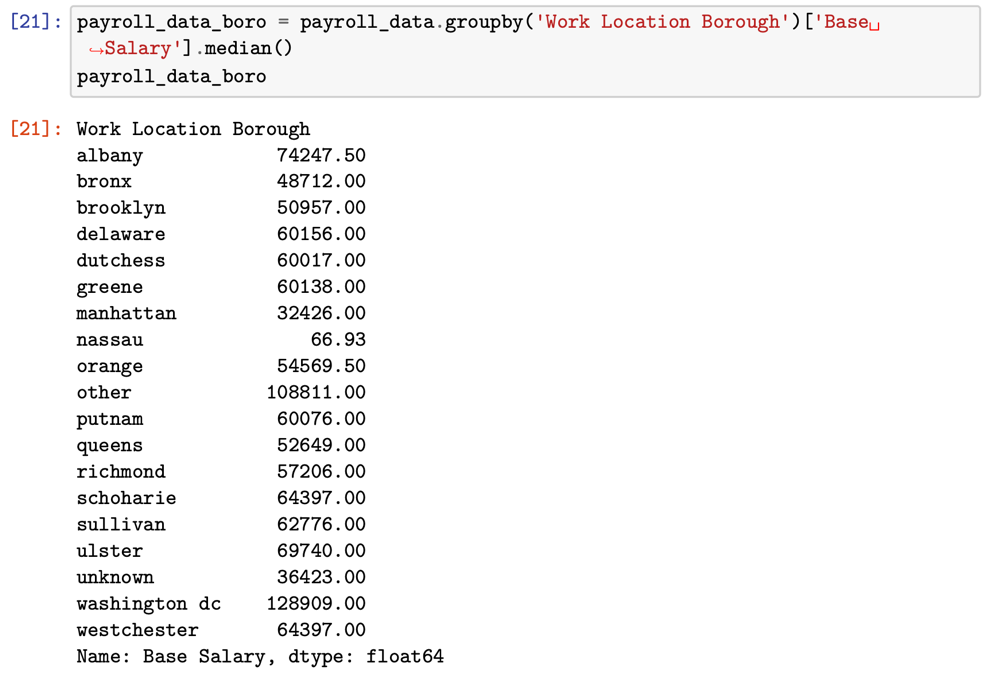
b. And, the following are the visualization of the highest amount of base salary from 2019 to 2021 (post COVID-19):



By comparing both bar plots, we can see how top8 job titles with highest base salary shifted before (2014-2018) and after (2019-2022) the COVID-19 pandemic. Jobs like School Safety Admin has gone up by a lot; Hardcore jobs like attorney still hold up their places in the top3.

**5.5 The median base salary in different borough**

a. we used the following codes to calculate the median of base salary in different borough from 2014 - 2022:

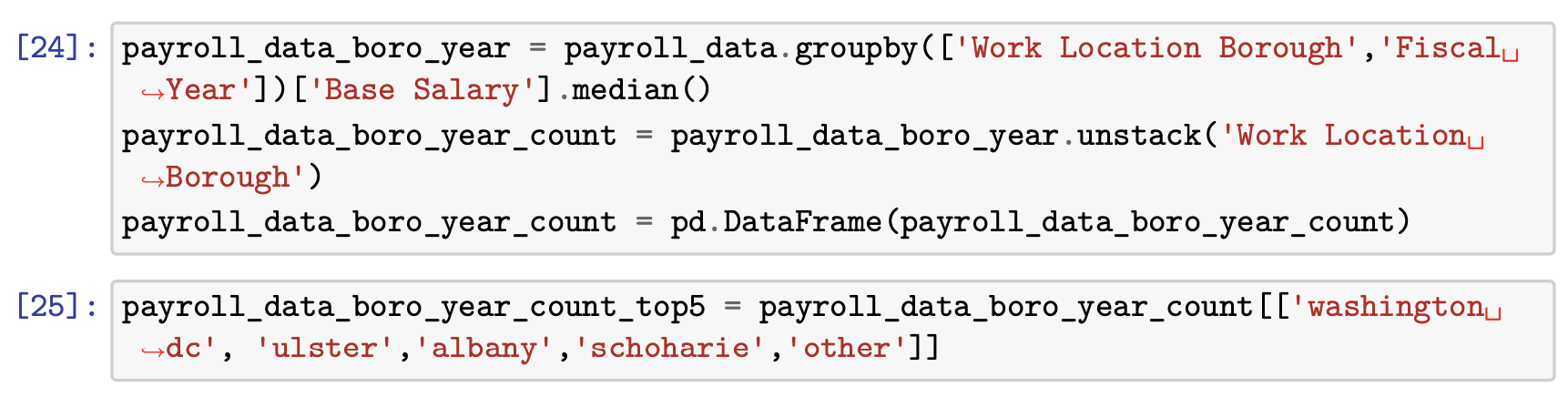


b. And, the following are the visualization of the median of base salary in different borough from 2014 - 2022:

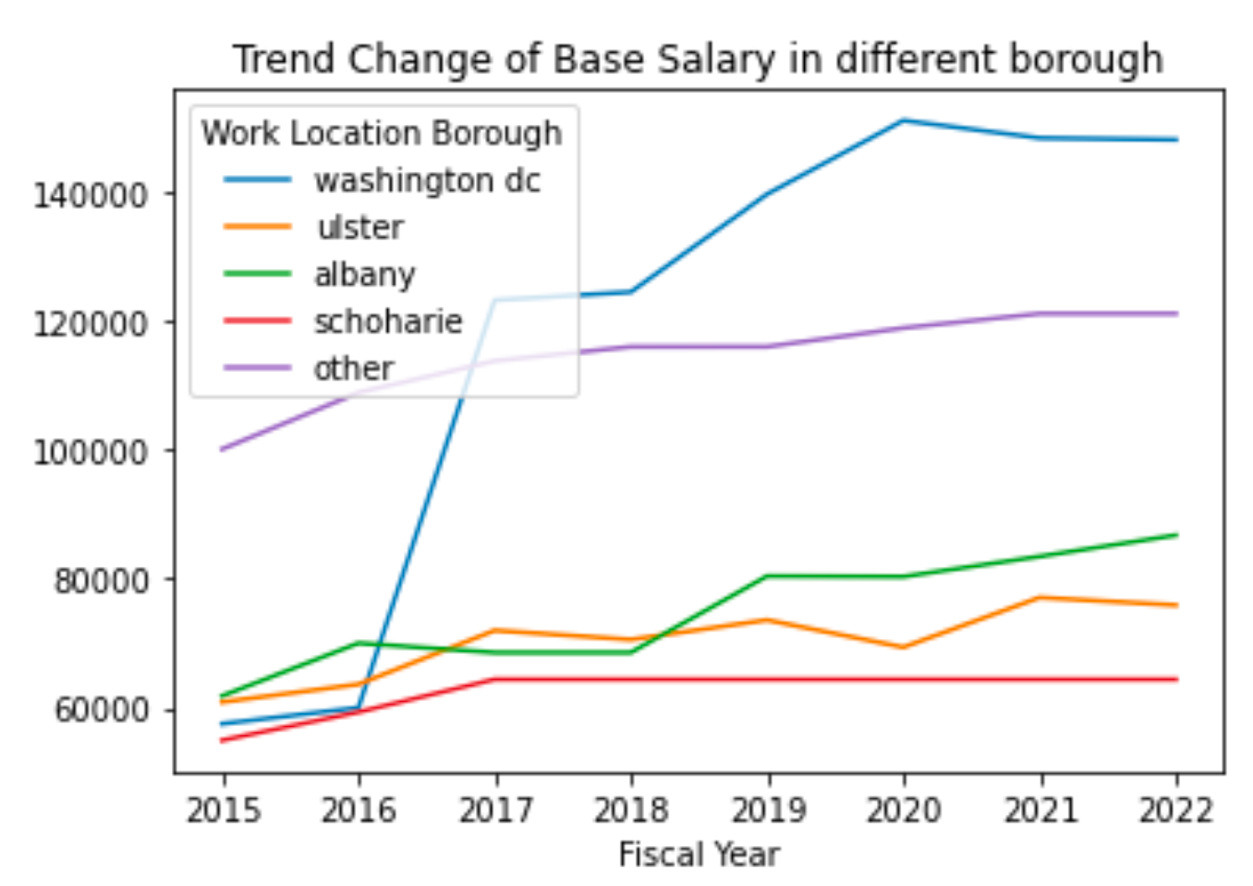


**5.6 Trend Change of the highest amount of in different borough**

a. we used the following codes to show the trend Change of the highest amount of in different borough:



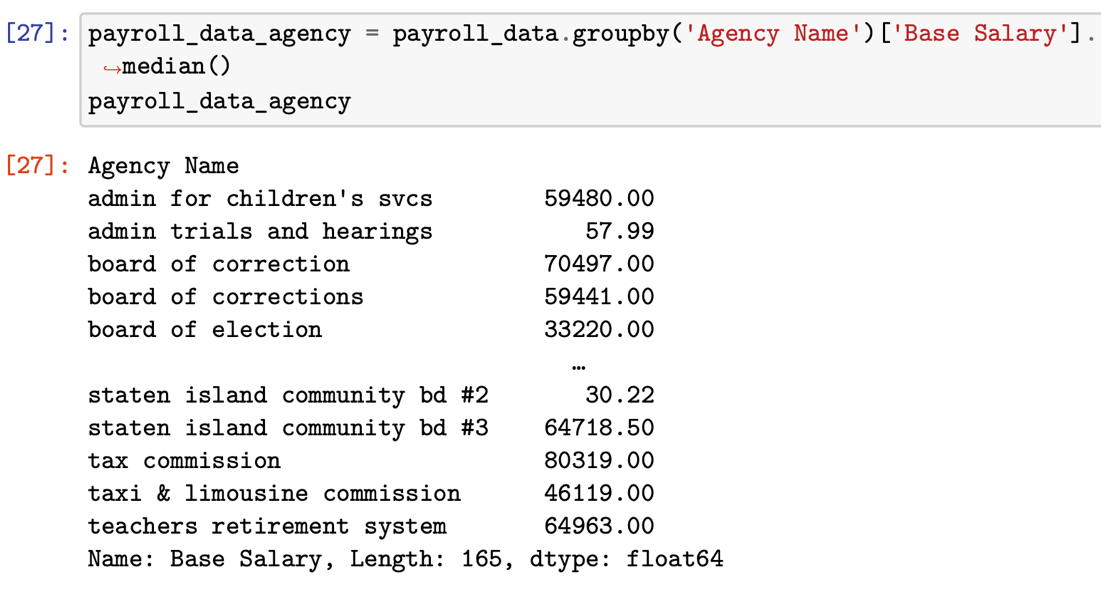
b. And, the following are the visualization of the trend Change of the highest amount of in different borough:



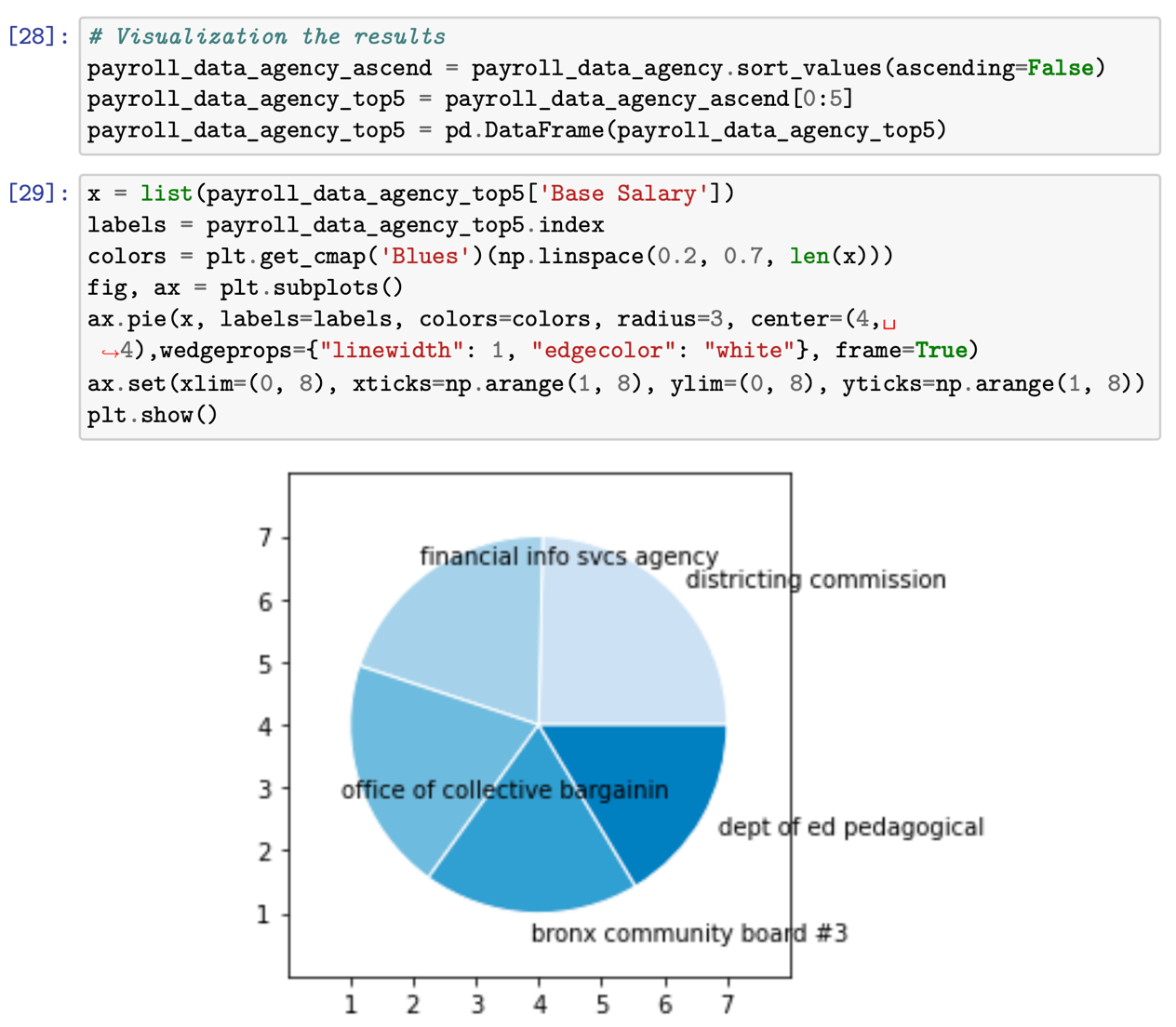
From the pie chart, we noticed that Washington DC is the borough which has the highest base salary in New York City. And from the line chart below, we found that Washington DC has a very high jump since year 2016. And always leads the place than other boroughs. Regards to the other 4 boroughs, seems like there were no major upwards and downwards from 2015 to 2022.

**5.7 Base Salary (median) amount in different agencies:**

a. we used the following codes to calculate the median of base salary in different agencies:

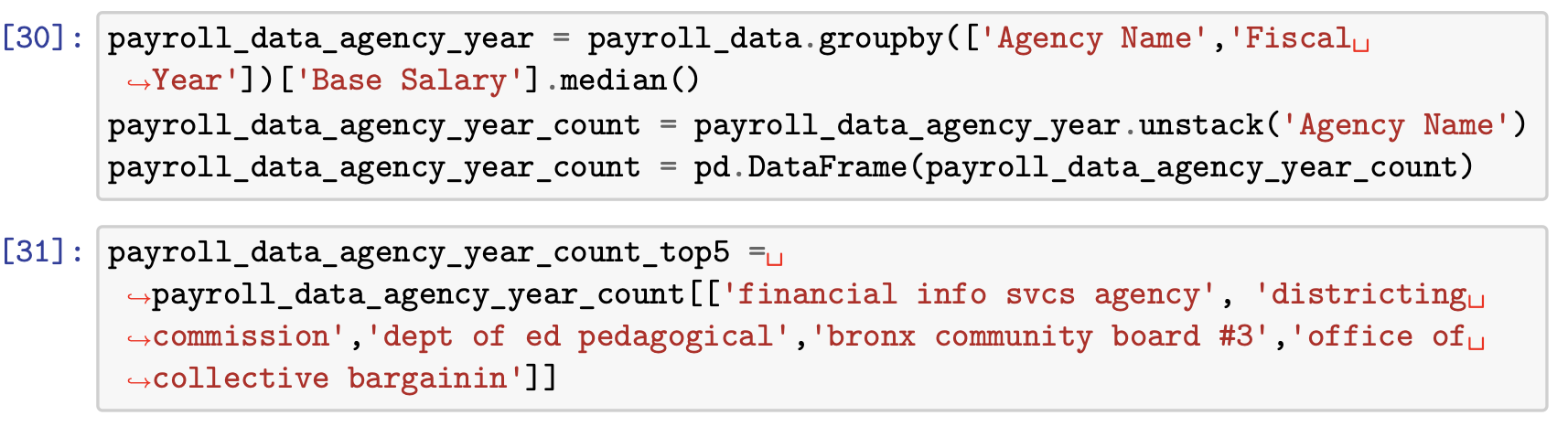


b. And, the following are the visualization of the base salary in different agencies:

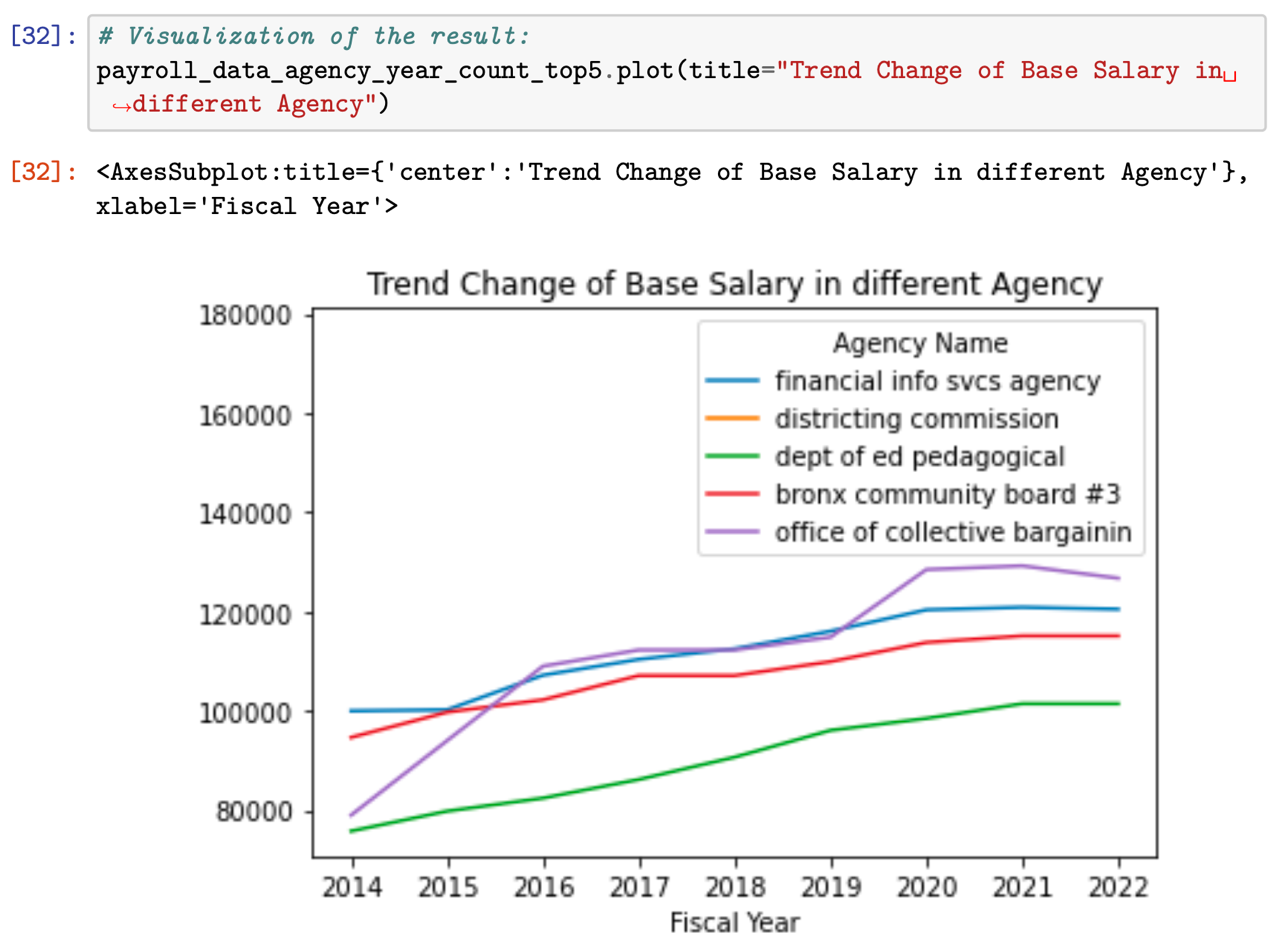


**5.8 Trend change of Base Salary (median) amount in different agencies:**

a. we used the following codes to show the trend of base salary in different agencies:



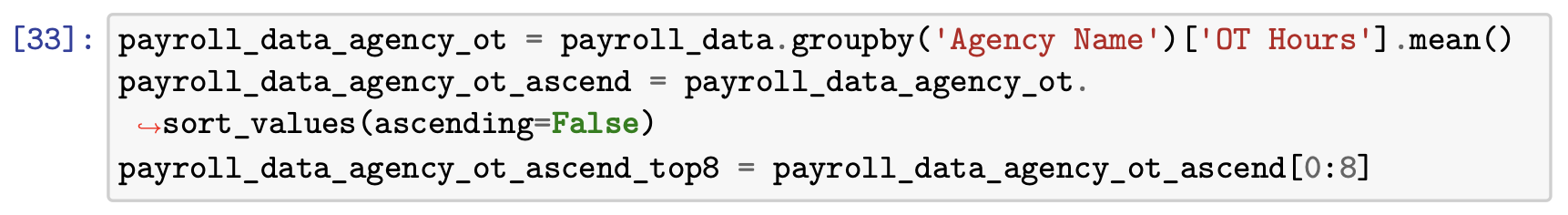
b. And, the following are the visualization of the trend:



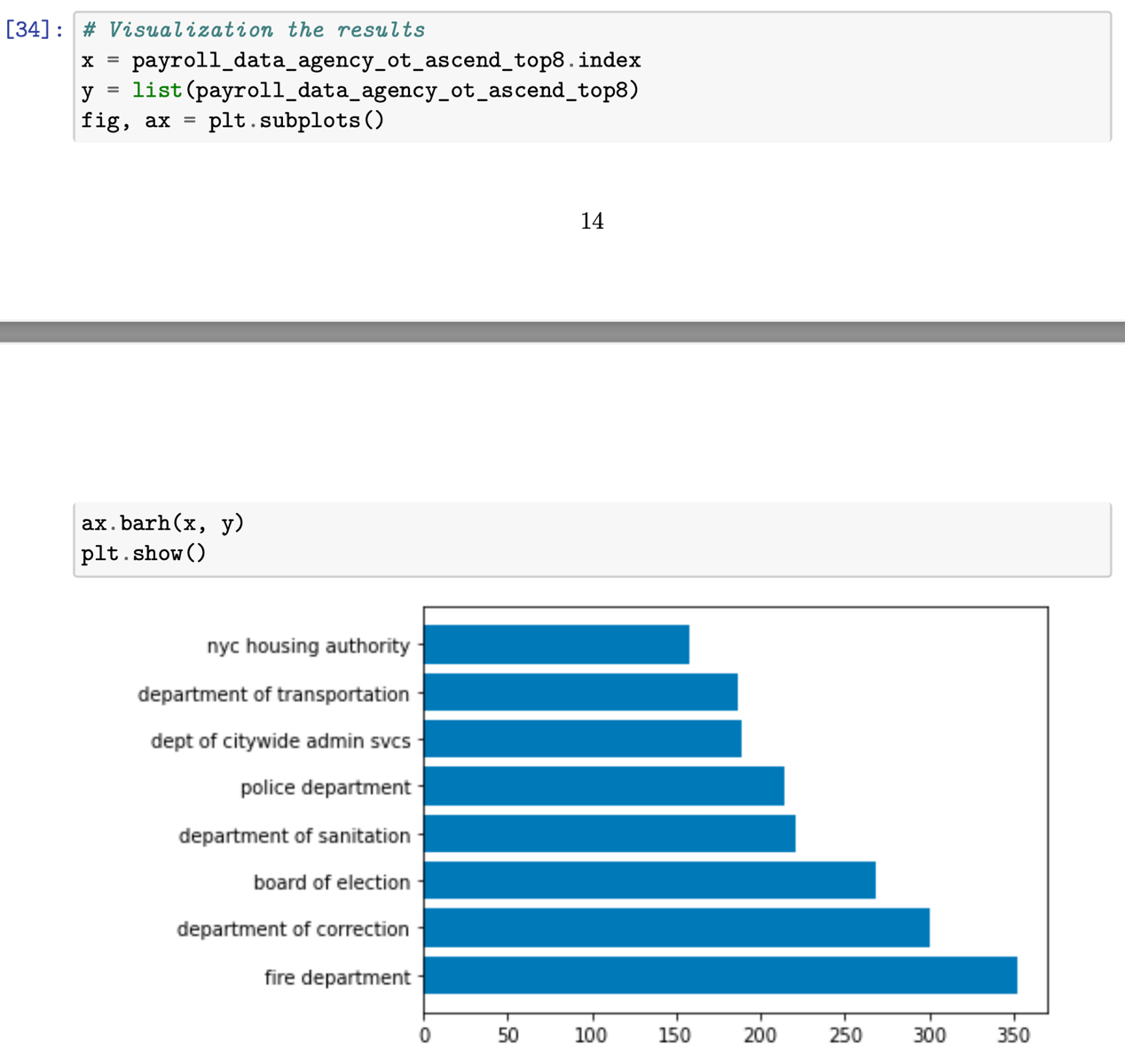
From the pie chart, we can see that top5 agencies which have the highest amount of base salary have almost the same amount of base salary from 2014 to 2022. However Districting Commission is the highest. And from the line chart, we found that the overall trend keeps going up from 2014 to 2022; however no major bump among top5 agencies. Good thing is the COVID-19 pandemic didn’t affect these agencies by a lot in regards of base salary they assigned to employees.

**5.9 Overtime Hours across each different agency:**

a. we used the following codes to show overtime hours across each different agency:

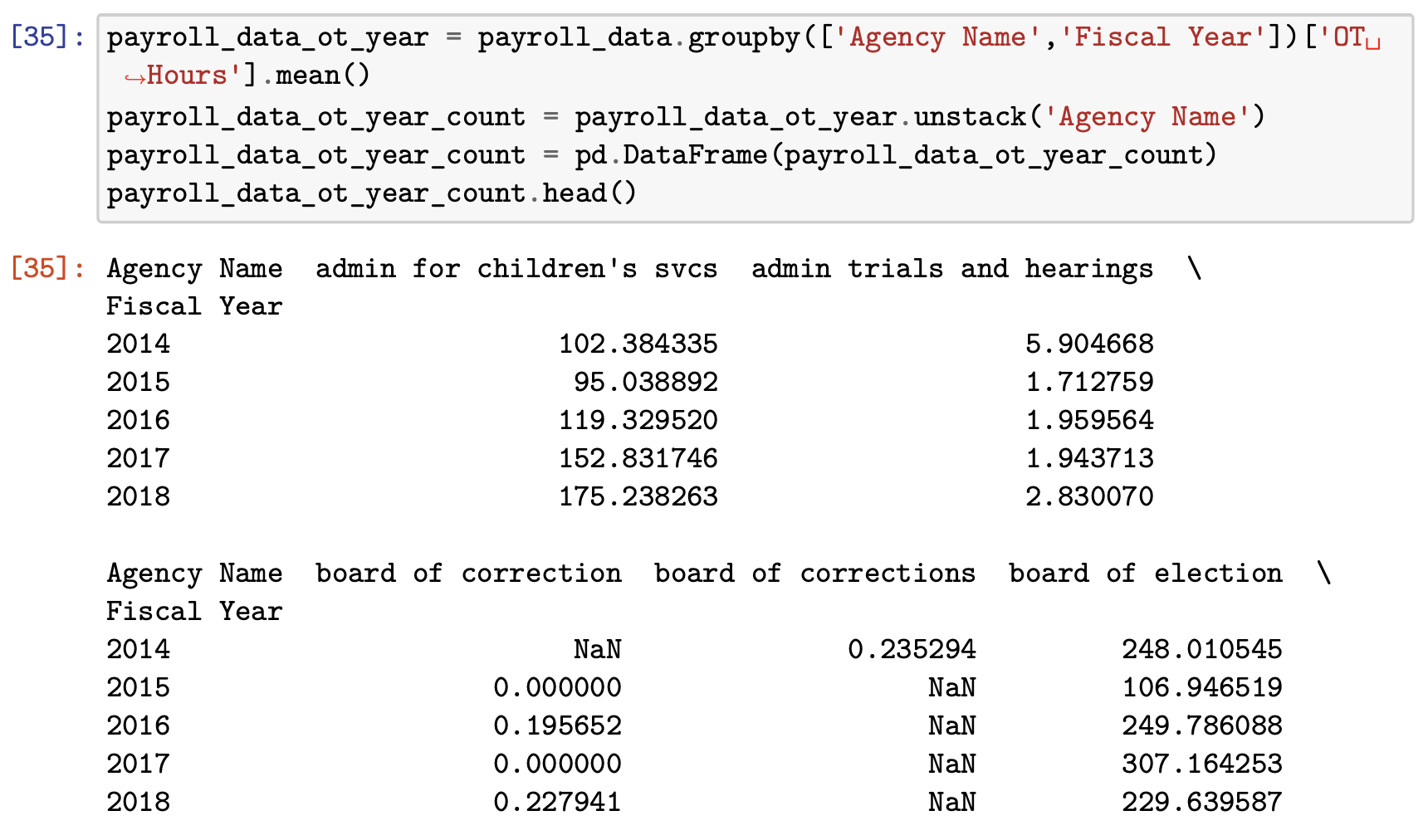


b. And, the following are the visualization:

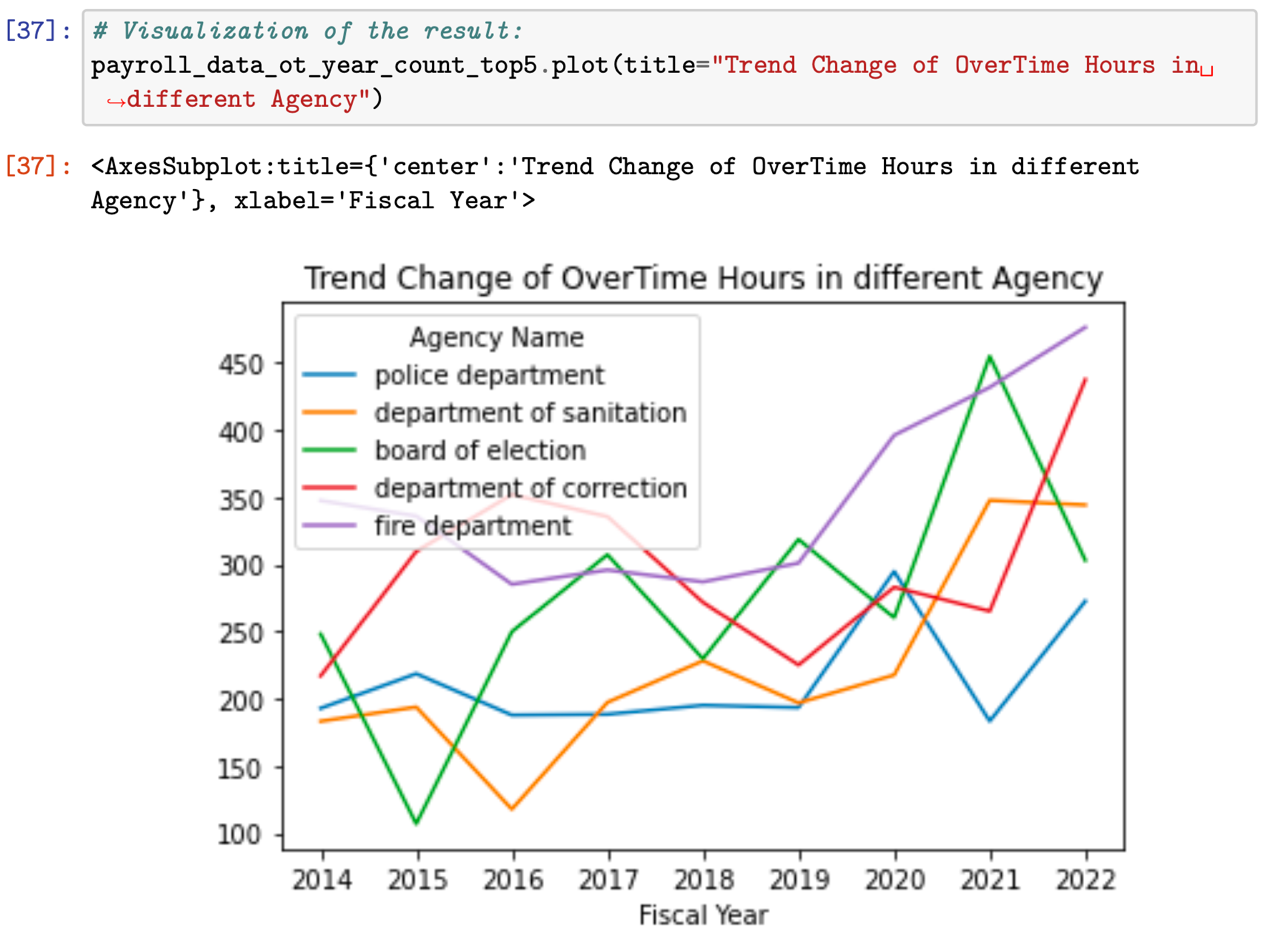


**5.10 Trend Change of Overtime Hours across each different agency:**

a. we used the following codes to show trend of overtime hours across each different agency:



b. And, the following are the visualization of the trend:



From the pie chart above, we noticed that the Fire Department has the highest amount of overtime hours which does not surprise us. However, when by looking at the line chart above, we noticed that most of these agencies’ overtime hours went up after the outbreak of COVID-19 in 2019. Especially for the Police Department, the overtime hours went up by lot in 2019. On the other hand, we notice that the Board of Election also had higher overtime hours over years. The reason behind this may because the election of President of United States.

**6. Conclusion**

* The average base salary was significantly higher just before the pandemic hit in 2019 than what they were in 2014. The median income had increased by $28,000 in 2014 to $43,000 in 2019 (53% increase). When entering the pandemic, the growth rate has slowed down by 16%. Post-pandemic these numbers have begun to decrease and showing a slow but steady improvement.
* Jobs like attorneys and school security managers were assigned with very high base salary among other job titles.
* Washington DC is tended to be the best city to work at, since it has the highest base salary and its trend is still going up and much higher than other cities.
* COVID -19 affected overtime hours on some hard-core agencies such as police departments caused by some safety issues around different cities during the quarantine.

**7. Source**

This data is input into the City's Personnel Management System by the respective user Agencies. Please refer to the attached URL for additional information about this dataset:

https://data.cityofnewyork.us/City-Government/Citywide-Payroll-Data-Fiscal-Year-/k397-673e