

# Big Data Analytic Applications Symposium - Spring 2020

---

Project Name: Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

Team Name: City Jobs Explorer

Team: Jiyuan Lu

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Abstract:

By comparing city employee payroll data between NYC and Chicago, we can gain insights as to which city has higher salary overall, what kind of city employee jobs tend to be better paid, and how each city's budget is being devoted overtime.

By comparing NYC payroll data and NYC job openings data over the years, we can observe the trend whether a department is hiring more employees, and whether city employees are getting better paid.

Finally, recommendations are given for perspective city employees on which city is more suitable for them to pursue their careers.

Platform(s) where the application runs:

NYU DUMBO cluster

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Motivation

Who are the users of this application?

The city government and job seekers who want to pursue city jobs.

Who will benefit from this application?

The city government and job seekers who want to pursue city jobs.

Why is this application important?

The city government will have a better idea on how the city's financial resources are allocated. And a more balanced budget distribution could be made for the next year.

The job seekers will get to know which city tends to have more job openings and better salary for their dream jobs. And they can take this information into consideration when deciding where to work.

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Goodness

What steps were taken to assess the ‘goodness’ of the analytic part of the application?

1. In the profiling step, check the number of records remaining after cleaning to ensure there are enough data for the analytics.
2. Check if the analytic results preserve the desired relationships. For example, the following relationship should hold:  $\text{Total budget increase percentage} = \text{Average salary increase percentage} * \text{Job positions increase percentage}$ .
3. Visualize the analytic results. For example, check if the distributions of the budget follow the same pattern for each city.

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Actuation/Remediation

What actuation or remediation actions can be performed?

The NYC government could spend more budget on police officers and cut down the number of unessential employees in the police department.

The Chicago government could re-evaluate whether it is necessary to spend most of its budget on the police department.

A notification email could be sent to the job seekers who subscribe to the NYC Jobs dataset when there is a new job posting that matches their search criteria.

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## All Data Sources

Name: Current Employee Names, Salaries, and Position Titles

Description: A listing of all current City of Chicago employees with information such as name, salary, department, and position title.

Size of data: 2.19 MB

Link to data: <https://data.cityofchicago.org/Administration-Finance/Current-Employee-Names-Salaries-and-Position-Title/xzkq-xp2w>

Name: Citywide Payroll Data (Fiscal Year)

Description: Each record represents statistics such as agency, name, work location, job title, and salary.

Size of data: 444 MB

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

Name: NYC Jobs

Description: Current job postings available on the City of New York's official job site.

Size of data: 13.1 MB

Link to data: <https://data.cityofnewyork.us/City-Government/NYC-Jobs/kpav-sd4t>

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Data Sample: Current Employee Names, Salaries, and Position Titles

DepartmentName	Salary	Title
FIRE	72510	PARAMEDIC
POLICE	93354	POLICE OFFICER
WATER MGMNT	132972	DEPUTY COMMISSIONER
POLICE	90024	POLICE OFFICER
POLICE	111444	SERGEANT
POLICE	103932	POLICE OFFICER (ASSIGNED AS DETECTIVE)
FIRE	107106	FIREFIGHTER (PER ARBITRATORS AWARD)-EMT
POLICE	95736	ASST DIR
POLICE	94524	POLICE OFFICER/EXPLSV DETECT K9 HNDLR
POLICE	48078	POLICE OFFICER
TRANSPORTN	67896	BRIDGE OPERATOR
POLICE	122472	SERGEANT
POLICE	90024	POLICE OFFICER
POLICE	84054	POLICE OFFICER
POLICE	84054	POLICE OFFICER
FIRE	89148	FIREFIGHTER-EMT
FIRE	92274	FIREFIGHTER-EMT
FINANCE	47160	ACCOUNTING TECHNICIAN II
LAW	143976	ADMINISTRATIVE DEPUTY
POLICE	111444	SERGEANT
POLICE	94122	POLICE OFFICER (ASSIGNED AS DETECTIVE)
WATER MGMNT	117072	CIVIL ENGINEER IV
POLICE	48078	POLICE OFFICER

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Data Sample: Citywide Payroll Data (Fiscal Year)

HEALTH	89980	DEPUTY DIRECTOR OF MEDICOLEGAL INVESTIGATIONS
SANITATION	159762	DEPUTY DIRECTOR OF MOTOR EQUIPMENT MAINTENANCE
SANITATION	148472	DEPUTY DIRECTOR OF MOTOR EQUIPMENT MAINTENANCE
SANITATION	148472	DEPUTY DIRECTOR OF MOTOR EQUIPMENT MAINTENANCE
SANITATION	148472	DEPUTY DIRECTOR OF MOTOR EQUIPMENT MAINTENANCE
SANITATION	148472	DEPUTY DIRECTOR OF MOTOR EQUIPMENT MAINTENANCE
HOUSING	201121	DEPUTY GENERAL COUNSEL
HOUSING	161136	DEPUTY GENERAL COUNSEL
HOUSING	161136	DEPUTY GENERAL COUNSEL
HOUSING	161136	DEPUTY GENERAL COUNSEL
HOUSING	156443	DEPUTY GENERAL COUNSEL
HEALTH	185629	DIRECTOR
HEALTH	137154	DIRECTOR
HEALTH	107424	DIRECTOR
HEALTH	96974	DIRECTOR
HEALTH	100205	DIRECTOR OF BIODETECTION AND RESPONSE
HEALTH	135584	DIRECTOR OF COMMUNITY ENGAGEMENT AND RESPONSE
HEALTH	101890	DIRECTOR OF COMMUNITY ENGAGEMENT AND RESPONSE
HEALTH	105338	DIRECTOR OF CRISIS ASSISTANCE AND TRAINING
HEALTH	211028	DIRECTOR OF FORENSIC BIOLOGY
HEALTH	130871	DIRECTOR OF HEALTH CARE PROGRAM PLANNING/ANALYSIS
HEALTH	93060	DIRECTOR OF HEALTH CARE PROGRAM PLANNING/ANALYSIS
HEALTH	92676	DIRECTOR OF HEALTH CARE PROGRAM PLANNING/ANALYSIS
HEALTH	119462	DIRECTOR OF INTERAGENCY PLANNING
SANITATION	213969	DIRECTOR OF MOTOR EQUIPMENT MAINTENANCE
SANITATION	188205	DIRECTOR OF MOTOR EQUIPMENT MAINTENANCE-NM
HOUSING	105000	DIRECTOR OF NEIGHBORHOOD CONSERVATION



# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

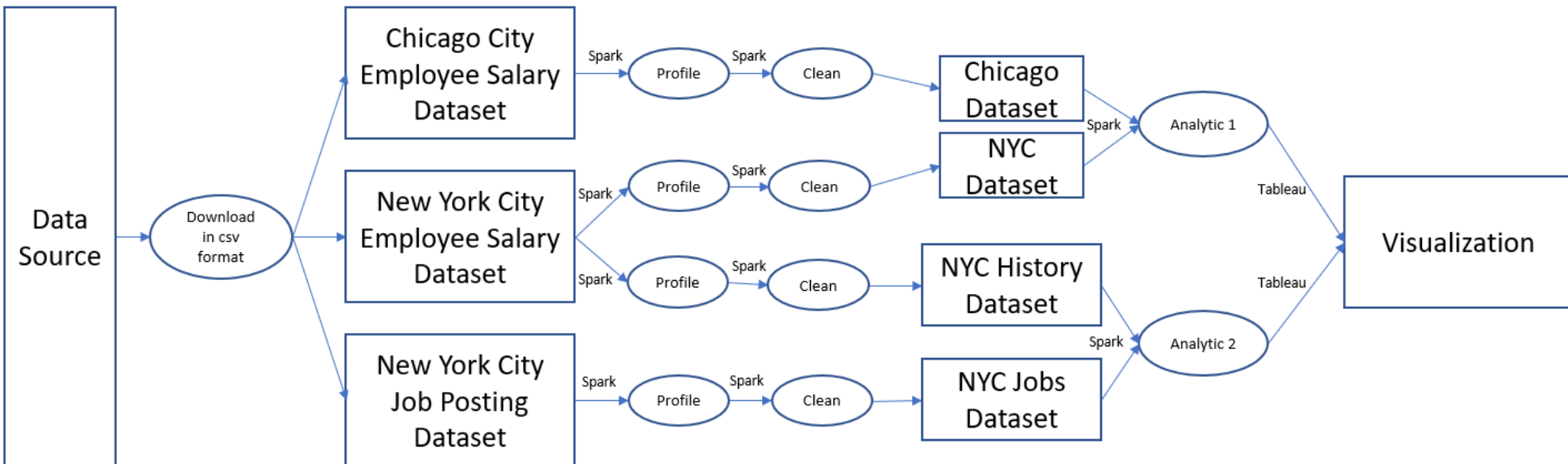
## Data Sample: NYC Jobs

Department	Title	Year	Salary
OFF OF PAYROLL ADMINISTRATION	Payroll Funding Analyst	2020	59,129.50
DEPARTMENT OF SANITATION	ARCHITECTURAL INTERN	2020	56,836.50
DEPT OF ENVIRONMENT PROTECTION	Civil Engineer I	2018	67,804.50
DEPARTMENT OF TRANSPORTATION	Civil Engineer 3	2018	103,050.00
DEPT OF ENVIRONMENT PROTECTION	PORTFOLIO MANAGER	2020	132,095.00
DEPARTMENT OF SANITATION	ELECTRICAL ENGINEER	2020	96,237.00
DEPT OF ENVIRONMENT PROTECTION	Dam Safety Coordinator	2017	77,210.50
DEPARTMENT OF SANITATION	Energy Manager	2020	110,117.00
NYC HOUSING AUTHORITY	PROJECT MANAGER	2019	81,478.00
DEPT OF ENVIRONMENT PROTECTION	Engineering Technician I	2019	40,630.50
DEPARTMENT OF BUSINESS SERV.	DATA ANALYST, BUSINESS OPERATIONS	2019	58,762.00
DEPT OF ENVIRONMENT PROTECTION	OFFICE ENGINEER	2019	80,526.50
NYC HOUSING AUTHORITY	CONFIDENTIAL INVESTIGATOR	2019	53,304.00
DEPT OF ENVIRONMENT PROTECTION	Financial Auditor	2018	58,571.50
ADMIN FOR CHILDREN'S SVCS	Director of Child Protective Audits	2017	81,739.50
DEPT OF HEALTH/MENTAL HYGIENE	Central Office Supervising Nurse, Bureau of School Health/SH Nursing Services & Prof Dev	2018	78,429.00
NYC HOUSING AUTHORITY	Analysis and Reporting Analyst	2019	94,889.00
TAXI & LIMOUSINE COMMISSION	Programs Data Manager	2019	66,050.00
DEPARTMENT OF BUSINESS SERV.	.NET DEVELOPER	2019	101,097.00
ADMIN FOR CHILDREN'S SVCS	Child Welfare Analyst	2016	76,941.50
DEPT OF ENVIRONMENT PROTECTION	STAFF ANALYST II	2016	64,384.00
NYC HOUSING AUTHORITY	Investigative Attorney	2020	69,078.50
DEPT OF ENVIRONMENT PROTECTION	Accountable Manager, CSO LTCP Program	2016	91,484.00
DEPT OF ENVIRONMENT PROTECTION	Information Technology Manager	2018	90,000.00
DEPT OF ENVIRONMENT PROTECTION	Minority Women Business Enterprise Program Analyst	2019	64,762.00
DEPT OF ENVIRONMENT PROTECTION	Associate Project Manager I	2017	77,210.50
DEPT OF ENVIRONMENT PROTECTION	Community Assistant	2019	36,267.50
ADMIN FOR CHILDREN'S SVCS	Policy Writer	2016	72,016.00

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Design Diagram



# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Code Challenge

### Combining and standardizing department names

```
chicago.registerTempTable("chicago")
val chicago_departments_renamed = sqlCtx.sql(""" SELECT CASE WHEN Department IN ("ADMIN HEARING",
"BOARD OF ELECTION", "BUDGET & MGMT", "BUILDINGS", "BUSINESS AFFAIRS", "CITY CLERK", "CITY COUNCIL",
"CULTURAL AFFAIRS", "FINANCE", "FIRE", "HEALTH", "HOUSING", "HUMAN RELATIONS", "LAW", "MAYOR'S OFFICE",
"POLICE", "TRANSPORTN", "TREASURER", "WATER MGMNT") THEN Department WHEN Department = "OEMC" THEN "EMERGENCY"
WHEN Department = "STREETS & SAN" THEN "SANITATION" ELSE "N/A" END AS DepartmentName, Title, Salary FROM chicago
WHERE Department != "N/A" """)

nyc.registerTempTable("nyc")
val nyc_departments_renamed = sqlCtx.sql(""" SELECT CASE WHEN Department IN ("BOARD OF ELECTION",
"CITY CLERK", "CITY COUNCIL", "CULTURAL AFFAIRS") THEN Department WHEN Department LIKE "%HEARNING%" THEN
"ADMIN TRAILS AND HEARINGS" WHEN Department LIKE "%BUDGET%" THEN "BUDGET & MGMT" WHEN Department LIKE
"%BUILDINGS%" THEN "BUILDINGS" WHEN Department LIKE "%BUSINESS%" THEN "BUSINESS AFFAIRS" WHEN
Department = "DEPARTMENT OF FINANCE" OR Department = "FINANCIAL INFO SVCS AGENCY" THEN "FINANCE" WHEN Department =
"FIRE DEPARTMENT" THEN "FIRE" WHEN Department LIKE "%HEALTH%" THEN "HEALTH" WHEN Department LIKE "%HOUSING%" THEN
"HOUSING" WHEN Department LIKE "%HUMAN%" OR Department LIKE "%RELATIONS%" THEN "HUMAN RELATIONS" WHEN Department LIKE "%LAW%"
THEN "LAW" WHEN Department = "OFFICE OF THE MAYOR" THEN "MAYOR'S OFFICE" WHEN Department LIKE "%EMERGENCY%"
THEN "EMERGENCY" WHEN Department = "POLICE DEPARTMENT" THEN "POLICE" WHEN Department LIKE "%SANITATION%"
THEN "SANITATION" WHEN Department LIKE "%TRANSPORTATION%" THEN "TRANSPORTN" WHEN Department LIKE "%ACTUARY%"
THEN "TREASURER" WHEN Department LIKE "%WATER%" THEN "WATER MGMNT" ELSE "N/A" END AS DepartmentName, Title, Salary FROM nyc """)
```

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Analytic Results

1. In 2019, the budget in NYC for full time city employees, full time police department employees, and full time police officers is 3.76, 3.06, and 1.85 times the budget for those of Chicago.
2. In 2019, among 22705 full time city employees in Chicago, 60.9% belong to the police department, among which 82.7% are police officers; among 104525 full time city employees in NYC, 48.5% belong to the police department, among which 48.7% are police officers.
3. In 2019, for Chicago, 60.2% of the city budget is spent on the police department, of which 79.2% is spent on police officers; for NYC, 48.8% of the city budget is spent on the police department, of which 48.0% is spent on police officers.
4. In 2019: The average salary for Chicago full time city employees is around 89.9K, with a standard deviation of about 20.7K, while the average salary for NYC full time city employees is around 73.4K, with a standard deviation of about 29.4K. The average salary for Chicago police staffs is around 88.8K, with a standard deviation of around 18.0K, while the average salary for NYC police staffs is around 74.1K, with a standard deviation of around 29.0K. The average salary for Chicago police officers is around 85.1K, with a standard deviation of about 11.8K, while the average salary for NYC police officers is around 73.1K, with a standard deviation of about 19.1K.
5. For NYC: The budget, average salary, and number of city jobs all increased each year. 29.2% in total budget, 16.3% in average salary, and 11.1% in the number of city jobs. However, the average salary of new city jobs does not vary much across the years.

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Insights

1. The first two ratios are in correspondence with the fact that NYC (8.4M) has approximately 3.1 times population than Chicago (2.7M). This means approximately the same amount of budget is spent on city employees and police staffs per capita in both cities. However, the last ratio shows that NYC spends less budget on police officers per capita (only 60% of that in Chicago). This shows the demand for police officers in NYC per capita is much less than that in Chicago, which in turn may indicate that NYC is safer and has lower crime rates than Chicago. In 2016, the violent crimes rates is 903.8 in Chicago, but only 585.8 in NYC.; the property crimes rates is 2946.2 in Chicago, but only 1518.7 in NYC.
2. Chicago is putting most of its financial resources in its police department, but this proportion is less than a half in New York. Within the police department, Chicago is primarily hiring police officers, while this proportion is also below a half in New York. This shows that the primary responsibility of the police department in Chicago may be to maintain order and control crimes, while the responsibility of the police department in NYC is more varied, e.g., park rangers.
3. In both cities, police and fire department employees make up for the majority of the city employees. And most of the city budget is spent on these two departments. This shows that the major functionality for both cities is to maintain social order.
4. Overall, the salary for full-time city employees in Chicago is higher than those in NYC, and is less variant. The same conclusion can also be made for the police department and for police officers. There is a tradeoff between salary and job opportunities for job seekers: Working in Chicago gets you better paid, but NYC offers more job opportunities. Also, higher salary for Chicago police officers may indicate that they potentially suffer from more risks than their counterparts in NYC due to higher crime rates.
5. Although there are more job openings and better salaries for full-time city jobs each year in NYC, the pay for new city employees stays approximately the same over the years. For job seekers, don't get frustrated by the low starting salary rate. You might get a pay rise soon!

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Obstacles

1. Although there are many city departments in common between Chicago and NYC, these departments are named differently. For example, “Fire” in the Chicago dataset but “Fire Department” in the NYC dataset. Therefore, the department names must be standardized before comparison.
2. When running Spark applications in cluster mode, there are few error messages. So first run the application on a subset of the data using the client mode to get more detailed error messages. After the bug is fixed, the application can be deployed on the cluster.
3. Tableau cannot identify .csv files in HDFS on DUMBO. In addition to saving files in .csv format, also create the corresponding Hive tables that Tableau can access.

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Summary

The population of New York City is approximately three times that of Chicago. This is reflected both in the budget spent on city employees and police staffs. However, NYC spends 40% less budget on police officers and has 30% fewer police officers per capita, and gives 14% less salary for police officers than Chicago. The evidence shows that the Chicago government puts more financial resources on maintaining social order and controlling crimes, and this corresponds to the fact that Chicago has a higher crime rate than NYC.

The police department and the fire department are the most important agencies in both cities. Most of the city budget is spent on the two departments. The majority of the city employees also belong to these two departments.

In terms of average salary, the best paid department in Chicago is the building department, and is the treasurer department in NYC. The worst paid department is the board of election for both cities.

For job seekers who wish to pursue a city job, especially as a police officer, Chicago provides better salary but higher risks, while NYC provides more job opportunities but less salary.

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

## Acknowledgements

Firstly, I appreciate the supports from NYU HPC staffs. They maintained the cluster and made it possible for me to store large data files and run heavy jobs on the cluster.

Secondly, I appreciate NYC Big Data Platform and Chicago Data Portal for generously sharing their data sources to the public.

Thirdly, I want to thank Tableau for providing me with the free trial license for the visualization.

Finally, I want to thank Professor McIntosh for giving me valuable advice and helping me solve technical problems throughout the research for this project.



# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

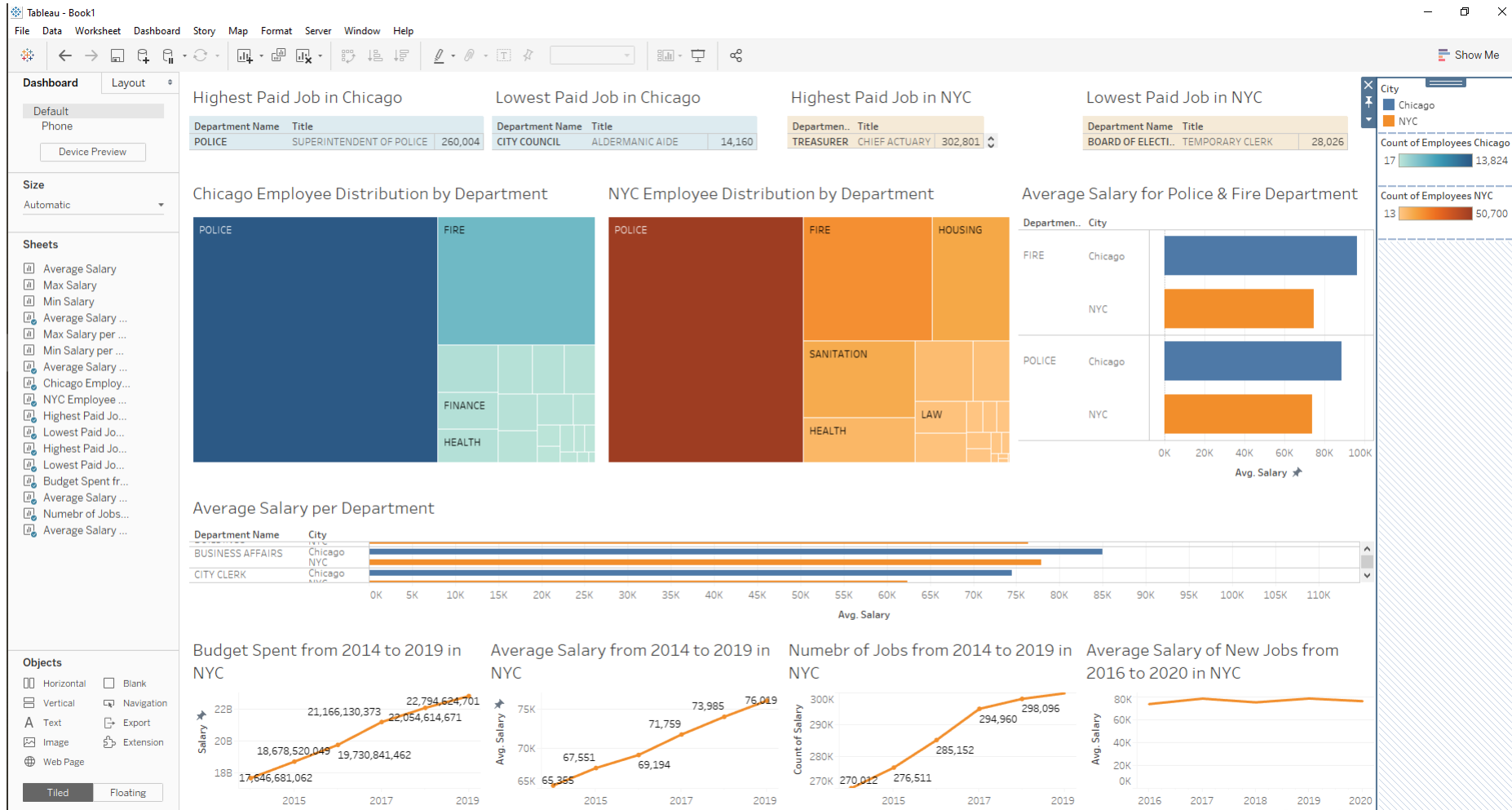
---

## References

1. Ester R. Fuchs. Mayors and money: Fiscal policy in New York and Chicago, University of Chicago Press, Chicago, IL, May 1992.
2. Walter Fogel & David Lewin. Wage Determination in the Public Sector, Industrial and Labor Relations Review, February 1974.
3. Current Employee Names, Salaries, and Position Titles. Chicago Data Portal.
4. Citywide Payroll Data (Fiscal Year). NYC OpenData.
5. NYC Jobs. NYC OpenData.

# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

## Demo!



# Who Earns More -- Comparing City Employee Payroll Data Between New York City and Chicago

---

Thank you!