NYC Payroll Data

Group 6
Nicole Cosmany
Ebrahim Azarisooreh

Questions

1. **Allocation:** How is the total NYC payroll budget allocated among departments and boroughs?
2. **Pay**: How does pay for a certain job title vary from year to year? Does the average pay for a certain job title vary across boroughs?
3. **Overtime**: How much of the total payroll spending is for overtime? Are certain departments/positions consistently spending more on overtime? Can we predict overtime?
4. **Duration of Employment:** How does pay rate/salary relate to employment duration? Do employees who are paid more/less have longer/shorter tenures than others? How does this vary across boroughs, departments, and job title?

Data preparation

- Special preprocessing of the data was implemented with a python script
- Hashtag patterns were removed from certain agency names, which were specific markers for unique years. Those patterns were removed and consolidated into one agency group.
- Jobs and agencies were normalized by uppercase to merge any disjoint data together.
- Hourly pay rates below 0 were excluded prior to computation
- One hot encoding

Tools Used

Python - Scikit-Learn, Pandas, Numpy, MatPlotLib, Seaborn

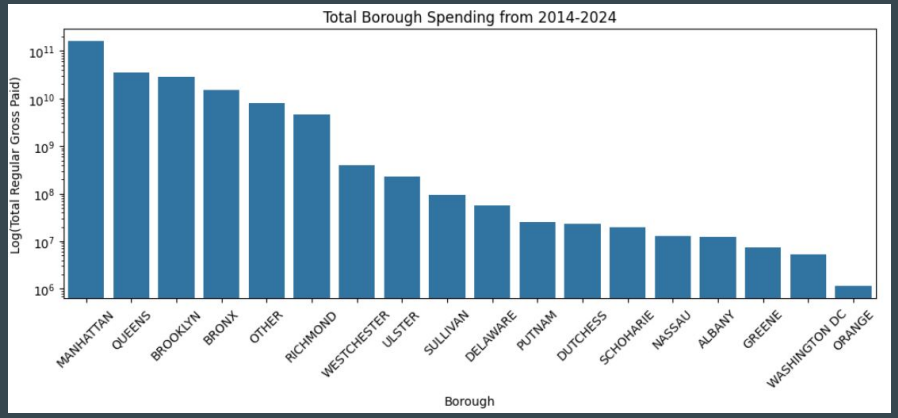
Github

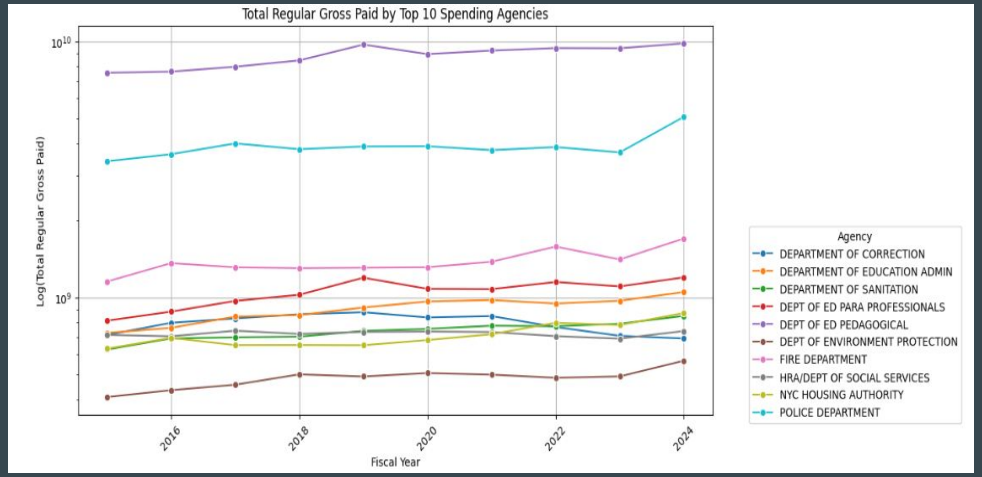
Google docs

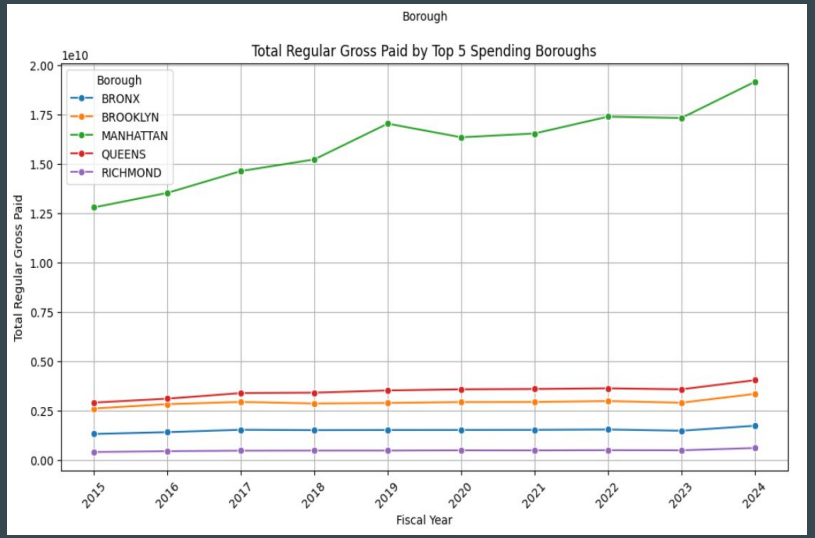
Analysis Methods

- Group by - sums, averages, etc.
- Standard deviation, coefficient of variation
- Linear regression models
- Visualizations: line plots, bar charts, scatter plots
- K-means clustering of job pay-grades: imputation of data to infer mean salaries

Results - Question 1



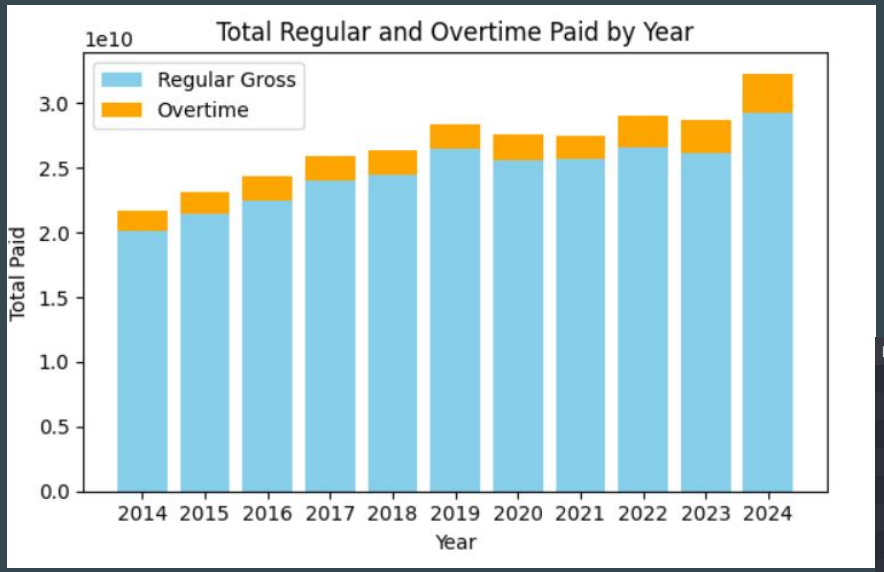


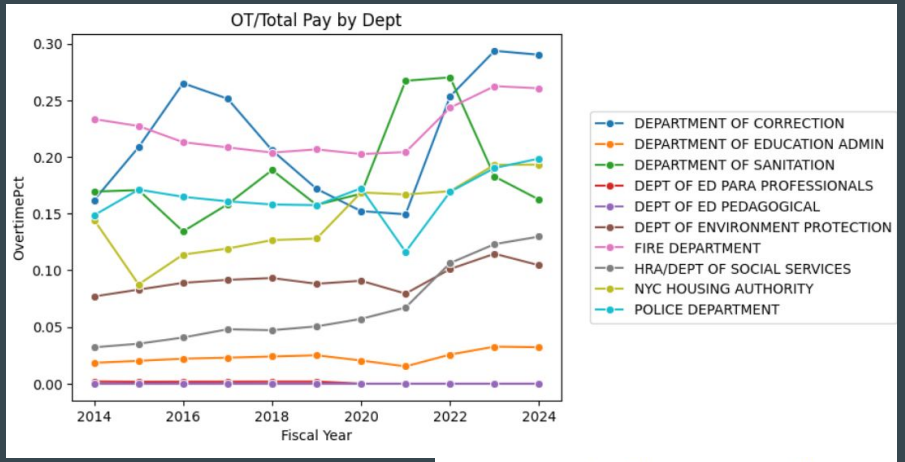


```
]:
# let's say I'm applying for a job in manhattan, and have 7 years of experience in the industry
y = [[0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 7, avg[11]]]

centroid, = kmeans.predict(y)
print(f'base salary rate = {kmeans.cluster_centers_[centroid][11]}')

# 14 years of experience in the Bronx?
y = [[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 14, avg[11]]]
centroid, = kmeans.predict(y)

print(f'base salary rate = {kmeans.cluster_centers_[centroid][11]}')
```

```
base salary rate = 51.192872673955065
base salary rate = 42.24896468571125
```

Results - Question 3





Fiscal Year	Regular Gross Paid	Total OT Paid	percent
2014	2.017930e+10	1.534109e+09	0.070653
2015	2.148923e+10	1.700476e+09	0.073329
2016	2.251816e+10	1.881360e+09	0.077106
2017	2.400997e+10	1.961936e+09	0.075541
2018	2.443846e+10	1.898494e+09	0.072085
2019	2.645986e+10	1.874010e+09	0.066140
2020	2.559189e+10	1.961704e+09	0.071196
2021	2.572134e+10	1.796557e+09	0.065287
2022	2.655873e+10	2.440510e+09	0.084158
2023	2.617739e+10	2.530125e+09	0.088135
2024	2.923852e+10	3.059251e+09	0.094720

Agency Name	mean	std	cv
DEPARTMENT OF CORRECTION	0.218476	0.054800	0.250828
DEPARTMENT OF EDUCATION ADMIN	0.023528	0.005330	0.226548
DEPARTMENT OF SANITATION	0.184419	0.043956	0.238351
DEPT OF ED PARA PROFESSIONALS	0.001007	0.000966	0.959181
DEPT OF ED PEDAGOGICAL	0.000000	0.000000	NaN
DEPT OF ENVIRONMENT PROTECTION	0.092009	0.011145	0.121132
FIRE DEPARTMENT	0.224152	0.022818	0.101798
HRA/DEPT OF SOCIAL SERVICES	0.067025	0.035604	0.531208
NYC HOUSING AUTHORITY	0.146488	0.034392	0.234777
POLICE DEPARTMENT	0.164322	0.021528	0.131010

Linear regression results

Y = # overtime hours worked

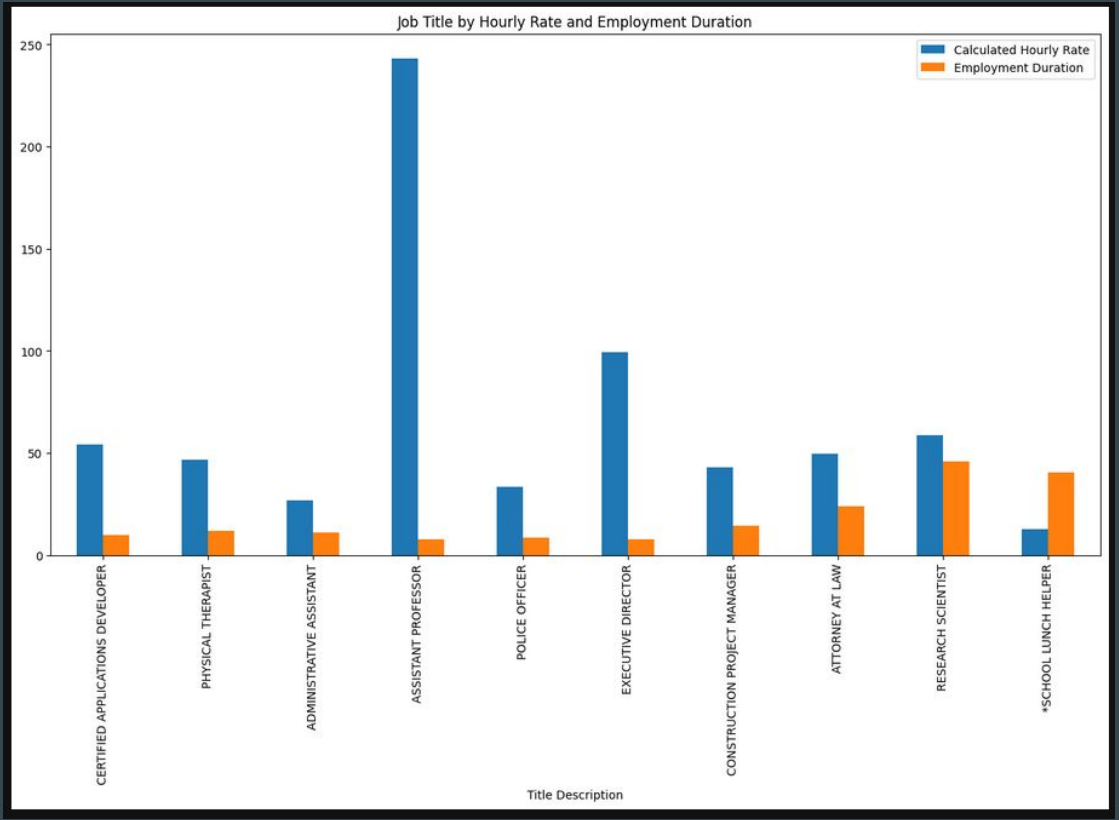
Predictor (X)	R squared
# employees	0.17
agency	0.97
# employees + agency	0.97

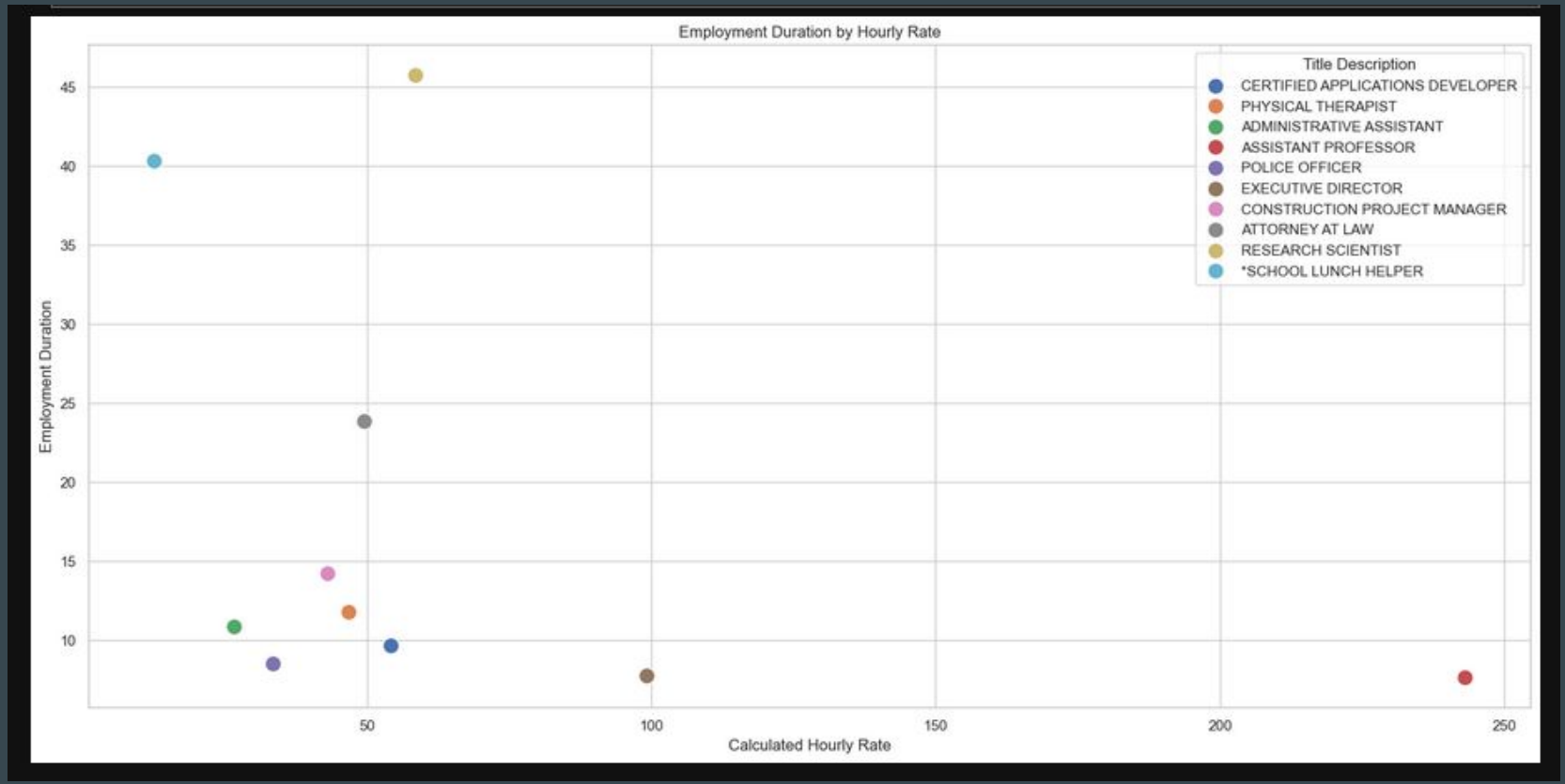
	Coefficient	Absolute Coefficient
Agency Name_POLICE DEPARTMENT	7.199632e+08	7.199632e+08
Agency Name_FIRE DEPARTMENT	3.749762e+08	3.749762e+08
Agency Name_DEPARTMENT OF CORRECTION	1.734702e+08	1.734702e+08
Agency Name_DEPARTMENT OF SANITATION	1.324518e+08	1.324518e+08
Agency Name_NYC HOUSING AUTHORITY	8.056942e+07	8.056942e+07

Top 5 Job Titles by OT Hours 2012-2022

	Title Description	Agency Name	OT Hours
6193	POLICE OFFICER	POLICE DEPARTMENT	61694804.28
5081	FIREFIGHTER	FIRE DEPARTMENT	44450310.38
3986	CORRECTION OFFICER	DEPARTMENT OF CORRECTION	34002181.01
6804	SANITATION WORKER	DEPARTMENT OF SANITATION	22011079.84
6064	P.O. DA DET GR3	POLICE DEPARTMENT	15683612.65

Results - Question 4




Employment Duration by Hourly Rate
Employment Duration
Calculated Hourly Rate
Title Description
CERTIFIED APPLICATIONS DEVELOPER
PHYSICAL THERAPIST
ADMINISTRATIVE ASSISTANT
ASSISTANT PROFESSOR
POLICE OFFICER
EXECUTIVE DIRECTOR
CONSTRUCTION PROJECT MANAGER
ATTORNEY AT LAW
RESEARCH SCIENTIST
*SCHOOL LUNCH HELPER
45
40
35
30
25
20
15
10
50
100
150
200
250

Application

Our results are useful for a wide range of audiences interested in NYC government spending and employee pay - including government planners, hiring managers, job seekers, and curious taxpayers.