Friendly Competition NYC Jobs Analysis

Geoffery Mullings

GeoffMullings@gmail.com

February 26, 2016

Executive Summary

The NYC Jobs data set was obtained from the NYC Open Data website. Out of all 45 agencies represented in the data set of open New York City government positions, the Department of Health and Mental Hygeine has the most openings: 1247. The Office of Administrative Trials and Hearings was overrepresented among the lowest paying positions, although the Office of Collective Bargaining houses the position offering the lowest compensation. On the other end the Department of Health and Mental Hygiene both dominated the highest paying positions and houses the top paying open job. When it comes to which jobs take the longest to fill the hypothesis is that the time it takes to fill a job is at least partially a function of the job's level, with higher level jobs requiring stronger qualifications taking longer to fill. The evidence found so far suggests that both the lowest level positions are actually proving harder to fill.

Importing the raw data and setting up shop.

The NYC Jobs data set was obtained from the NYC Open Data website on Feburary 22nd, 2016. The latest data set can be found here: https://data.cityofnewyork.us/City-Government/NYC-Jobs/kpav-sd4t

```
library("plyr")
library("lattice")
NYC_Jobs <- read.csv("~/Documents/rWD/NYC_Jobs.csv")
```

Question 1: Which agency has the most openings?

```
count(NYC_Jobs, "Agency")
```

```
##                              Agency freq
## 1       ADMIN FOR CHILDREN'S SVCS  167
## 2       ADMIN TRIALS AND HEARINGS   33
## 3   BUSINESS INTEGRITY COMMISSION    6
## 4    CIVILIAN COMPLAINT REVIEW BD   22
## 5     CONFLICTS OF INTEREST BOARD    4
## 6                CONSUMER AFFAIRS   28
## 7                CULTURAL AFFAIRS    2
## 8        DEPARTMENT FOR THE AGING   24
## 9         DEPARTMENT OF BUILDINGS  175
## 10   DEPARTMENT OF BUSINESS SERV.   58
```

```
## 11    DEPARTMENT OF CITY PLANNING    6
## 12       DEPARTMENT OF CORRECTION  107
## 13          DEPARTMENT OF FINANCE   86
## 14    DEPARTMENT OF INVESTIGATION  148
## 15        DEPARTMENT OF PROBATION    2
## 16       DEPARTMENT OF SANITATION  110
## 17   DEPARTMENT OF TRANSPORTATION  122
## 18    DEPT OF CITYWIDE ADMIN SVCS   38
## 19  DEPT OF DESIGN & CONSTRUCTION  278
## 20 DEPT OF ENVIRONMENT PROTECTION  398
## 21  DEPT OF HEALTH/MENTAL HYGIENE 1247
## 22   DEPT OF INFO TECH & TELECOMM  227
## 23     DEPT OF PARKS & RECREATION   82
## 24 DEPT OF RECORDS & INFO SERVICE    4
## 25  DEPT OF YOUTH & COMM DEV SRVS   32
## 26     DEPT. OF HOMELESS SERVICES   32
## 27 DISTRICT ATTORNEY KINGS COUNTY    4
## 28 DISTRICT ATTORNEY RICHMOND COU   10
## 29     FINANCIAL INFO SVCS AGENCY   27
## 30                FIRE DEPARTMENT   58
## 31 HOUSING PRESERVATION & DVLPMNT  146
## 32    HRA/DEPT OF SOCIAL SERVICES   71
## 33        HUMAN RIGHTS COMMISSION   20
## 34                  LAW DEPARTMENT   38
## 35   NYC EMPLOYEES RETIREMENT SYS    8
## 36          NYC HOUSING AUTHORITY   61
## 37        NYC POLICE PENSION FUND    4
## 38  OFF OF PAYROLL ADMINISTRATION    9
## 39 OFFICE OF COLLECTIVE BARGAININ    2
## 40 OFFICE OF EMERGENCY MANAGEMENT   16
## 41      OFFICE OF LABOR RELATIONS    2
## 42  OFFICE OF MANAGEMENT & BUDGET    6
## 43          OFFICE OF THE ACTUARY    1
## 44      OFFICE OF THE COMPTROLLER   14
## 45    TAXI & LIMOUSINE COMMISSION   28
```

It's clear that, with 292 openings, the Department of Health and Mental Hygeine has the most job openings. But which agencies have the highest and lowest paying positions? We'll need to do some data munging to attain that answer.

Data Munging

First, let's assign the most useful columns for analysis of our raw data to a variable. We'll call it "ConJobs" to stand for Consolidated Jobs.

After that, we'll address the issue of how salaries are recorded in the raw data. Rather than look at the minimum and maximum salary for each position we'll take an average of the two to use for our comparison. We'll call the variable holding that column of data "AvgSal" to stand for Average Salary, and add it to our ConJobs variable.

```
ConJobs = NYC_Jobs[c("Level","Agency","Salary.Range.From","Salary.Range
.To","Salary.Frequency","Posting.Date","Hours.Shift","Work.Location.1")
]
colnames(ConJobs)
```

```
## [1] "Level"             "Agency"            "Salary.Range.From"
## [4] "Salary.Range.To"   "Salary.Frequency"  "Posting.Date"
## [7] "Hours.Shift"       "Work.Location.1"
```

```
AvgSal = rowMeans(ConJobs[c("Salary.Range.From", "Salary.Range.To")], n
a.rm=TRUE)
head(AvgSal)
```

```
## [1] 105000.0     15.5     15.5  71780.5  71780.5  48549.0
```

```
ConJobs["AvgSal"] = AvgSal
```

For due dilligence, let's make a new variable that reflects ConJobs sorted from lowest to highest average salary. We'll call it ConJobsOrd to stand for Consolidated Jobs Ordered.

```
ConJobsOrd = arrange(ConJobs, AvgSal)
```

A bit more munging is now clearly necessary. Unfortunately analyzing the City's salary data is confounded by a mix of annual and hourly wages in the average salary column. We'll transform these observations into annual data.

Using the Salary.Frequency column we'll adjust hourly wages by putting them into their own variable, multiplying them by 2080 (an estimate of how many hours per year an employee would work), and replacing the original observations with the transformed ones. A similar operation will be done on the daily wages, but they will be multiplied by 260 instead (an estimate of how many days per year an employee would work). We'll accept that both estimates can't accurately account for missed days of work due to sickness or vacation.

```
levels(ConJobs$Salary.Frequency)
```

```
## [1] "Annual" "Daily"  "Hourly"
```

Salary jobs will have their own variable, ConJobsSal. For good measure hourly jobs will also have their own variable, ConJobsHR, as will daily jobs, ConJobsDay.

```
ConJobsSal = subset(ConJobs, ConJobs$Salary.Frequency == "Annual")
ConJobsHR = subset(ConJobs, ConJobs$Salary.Frequency == "Hourly")
ConJobsDay = subset(ConJobs, ConJobs$Salary.Frequency == "Daily")

#Transforming the data now.

ConJobsHR$AvgSal = ConJobsHR$AvgSal*2080
ConJobsDay$AvgSal = ConJobsDay$AvgSal*260

#Putting it all back together now.
```

```
ConJobs = rbind(ConJobsDay, ConJobsHR, ConJobsSal)
```

Data Analysis

Question 2: Which agencies have the highest and lowest paying positions?

Determining which agencies have the highest and lowest paying positions may be more complicated than just pulling the tail and head end of our ordered data. We should know about the distribution of pay among the open, salaried positions, and use descriptive statistical knowledge to identify which positions are truly the lowest and highest paying.

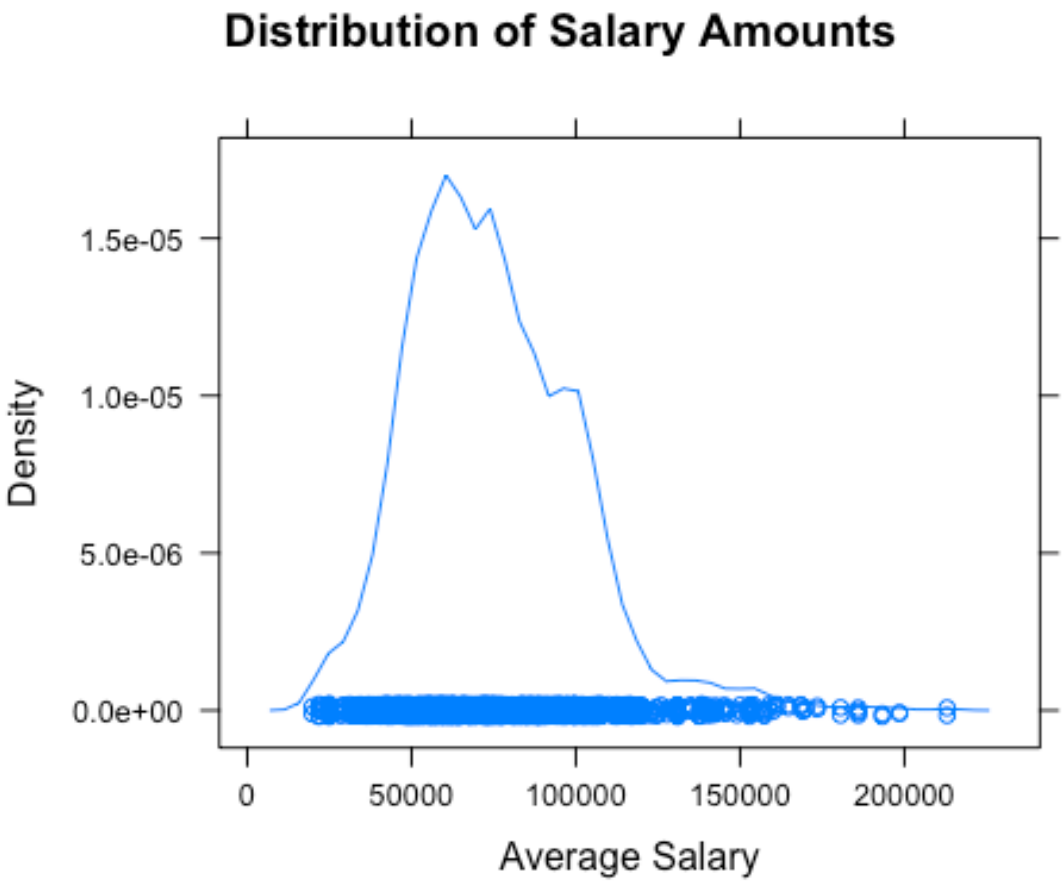
```
summary(ConJobs$AvgSal)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   19760   55920   72530   74630   89500  213000
```

```
sd(ConJobs$AvgSal)
```

```
## [1] 25428.24
```

Based on the mean and standard deviation of the subset of data, we can gather that any position with a salary below $23773.52 is below 97.5% of the data, arguably lowest paying.



On the other hand we could have used an exogeneous determinant of "lowest paying" position", such as NYC's poverty line measure. That may be more externally applicable, but it may also avoid the spirit of relativity in the question.

Let's build a table of all agencies in the lowest 2.5% of the average salary distribution.

```
sort(table(ConJobs[AvgSal <= 23773.52, "Agency"]))
```

```
##
##      ADMIN FOR CHILDREN'S SVCS  BUSINESS INTEGRITY COMMISSION
##                              0                              0
##   CIVILIAN COMPLAINT REVIEW BD    CONFLICTS OF INTEREST BOARD
##                              0                              0
##               CONSUMER AFFAIRS               CULTURAL AFFAIRS
##                              0                              0
##       DEPARTMENT FOR THE AGING        DEPARTMENT OF BUILDINGS
##                              0                              0
##   DEPARTMENT OF BUSINESS SERV.    DEPARTMENT OF CITY PLANNING
##                              0                              0
##       DEPARTMENT OF CORRECTION          DEPARTMENT OF FINANCE
##                              0                              0
##    DEPARTMENT OF INVESTIGATION        DEPARTMENT OF PROBATION
```

```
##                                0                                0
##       DEPARTMENT OF SANITATION   DEPARTMENT OF TRANSPORTATION
##                                0                                0
##    DEPT OF CITYWIDE ADMIN SVCS  DEPT OF DESIGN & CONSTRUCTION
##                                0                                0
## DEPT OF ENVIRONMENT PROTECTION  DEPT OF HEALTH/MENTAL HYGIENE
##                                0                                0
##   DEPT OF INFO TECH & TELECOMM     DEPT OF PARKS & RECREATION
##                                0                                0
## DEPT OF RECORDS & INFO SERVICE  DEPT OF YOUTH & COMM DEV SRVS
##                                0                                0
##     DEPT. OF HOMELESS SERVICES DISTRICT ATTORNEY KINGS COUNTY
##                                0                                0
## DISTRICT ATTORNEY RICHMOND COU     FINANCIAL INFO SVCS AGENCY
##                                0                                0
##                FIRE DEPARTMENT HOUSING PRESERVATION & DVLPMNT
##                                0                                0
##    HRA/DEPT OF SOCIAL SERVICES        HUMAN RIGHTS COMMISSION
##                                0                                0
##                 LAW DEPARTMENT   NYC EMPLOYEES RETIREMENT SYS
##                                0                                0
##          NYC HOUSING AUTHORITY        NYC POLICE PENSION FUND
##                                0                                0
##  OFF OF PAYROLL ADMINISTRATION OFFICE OF EMERGENCY MANAGEMENT
##                                0                                0
##      OFFICE OF LABOR RELATIONS  OFFICE OF MANAGEMENT & BUDGET
##                                0                                0
##          OFFICE OF THE ACTUARY      OFFICE OF THE COMPTROLLER
##                                0                                0
##    TAXI & LIMOUSINE COMMISSION OFFICE OF COLLECTIVE BARGAININ
##                                0                                2
##      ADMIN TRIALS AND HEARINGS
##                                8
```

It's clear that the Office of Administrative Trials and Hearings and the Office of Collective Bargaining make up the most common agencies among the lowest paying jobs. Collective Bargaining is home to the lowest paid position.

```
min(ConJobs$AvgSal)
```

```
## [1] 19760
```

```
ConJobsOrd = (ConJobs[order(AvgSal), ])
ConJobsOrd[1, ]
```

```
##     Level                        Agency Salary.Range.From Salary.Ra
nge.To
## 177    01 OFFICE OF COLLECTIVE BARGAININ                 9
10
##     Salary.Frequency        Posting.Date
## 177           Hourly 10/09/2014 00:00:00
```

```
##
Hours.Shift
## 177 Hours:  17 hours per week when school is in session, or 35 hours
per week during the summer vacation.                          Shift:
Weekdays/minimum three (3) mornings per week.
##     Work.Location.1 AvgSal
## 177                   19760
```

But which agencies offer the lowest average salary? Let's take the two most common agencies in the bottom and get their overall average salaries.

```
ConJobsOATH = subset(ConJobs, Agency =="ADMIN TRIALS AND HEARINGS")
mean(ConJobsOATH$AvgSal)
```

```
## [1] 51542.61
```

```
ConJobsOCB = subset(ConJobs, Agency =="OFFICE OF COLLECTIVE BARGAININ")
mean(ConJobsOCB$AvgSal)
```

```
## [1] 19760
```

Here it's made obvious that the Office Of Collective Bargaining is put at a disadvantage by having only two jobs to contribute to its average.

On the other end, where can you find the highest paying jobs and how much do those agencies pay on average?

Let's repeat the process on the other tail of the distribution: which agencies have the most open positions paying more than 97.5% of the other open positions?

```
sort(table(ConJobs[AvgSal >= 125486.5, "Agency"]))
```

```
##
##      ADMIN TRIALS AND HEARINGS  BUSINESS INTEGRITY COMMISSION
##                              0                              0
##    CONFLICTS OF INTEREST BOARD              CULTURAL AFFAIRS
##                              0                              0
##       DEPARTMENT FOR THE AGING        DEPARTMENT OF BUILDINGS
##                              0                              0
##    DEPARTMENT OF CITY PLANNING          DEPARTMENT OF FINANCE
##                              0                              0
##    DEPARTMENT OF INVESTIGATION        DEPARTMENT OF PROBATION
##                              0                              0
##  DEPT OF DESIGN & CONSTRUCTION     DEPT OF PARKS & RECREATION
##                              0                              0
## DEPT OF RECORDS & INFO SERVICE  DEPT OF YOUTH & COMM DEV SRVS
##                              0                              0
## DISTRICT ATTORNEY KINGS COUNTY DISTRICT ATTORNEY RICHMOND COU
##                              0                              0
##     FINANCIAL INFO SVCS AGENCY    HRA/DEPT OF SOCIAL SERVICES
##                              0                              0
##        HUMAN RIGHTS COMMISSION                 LAW DEPARTMENT
```

```
##                                0                                0
##   NYC EMPLOYEES RETIREMENT SYS          NYC HOUSING AUTHORITY
##                                0                                0
##        NYC POLICE PENSION FUND  OFF OF PAYROLL ADMINISTRATION
##                                0                                0
## OFFICE OF COLLECTIVE BARGAININ OFFICE OF EMERGENCY MANAGEMENT
##                                0                                0
##      OFFICE OF LABOR RELATIONS  OFFICE OF MANAGEMENT & BUDGET
##                                0                                0
##          OFFICE OF THE ACTUARY    TAXI & LIMOUSINE COMMISSION
##                                0                                0
##   CIVILIAN COMPLAINT REVIEW BD   DEPARTMENT OF TRANSPORTATION
##                                2                                2
##    DEPT OF CITYWIDE ADMIN SVCS HOUSING PRESERVATION & DVLPMNT
##                                2                                2
##      OFFICE OF THE COMPTROLLER      ADMIN FOR CHILDREN'S SVCS
##                                2                                4
##   DEPT OF INFO TECH & TELECOMM                FIRE DEPARTMENT
##                                4                                4
##               CONSUMER AFFAIRS     DEPT. OF HOMELESS SERVICES
##                                6                                6
## DEPT OF ENVIRONMENT PROTECTION   DEPARTMENT OF BUSINESS SERV.
##                               10                               12
##       DEPARTMENT OF CORRECTION       DEPARTMENT OF SANITATION
##                               12                               14
##  DEPT OF HEALTH/MENTAL HYGIENE
##                               52
```

It seems that the Department of Health and Mental Hygiene provides a lion's share of the highest paying positions in the city. The Department of Environmental Protection is a close second although it has the highest salaried position at $198,518.

```
max(ConJobs$AvgSal)
```

```
## [1] 213000
```

```
ConJobsOrd[3963, ]
```

```
##      Level                        Agency Salary.Range.From Salary.Ra
nge.To
## 1853    M7 DEPT OF HEALTH/MENTAL HYGIENE           213000
213000
##      Salary.Frequency        Posting.Date Hours.Shift Work.Location.
1
## 1853           Annual 10/30/2015 00:00:00
##      AvgSal
## 1853 213000
```

Healthcare is one of the fastest growing and most profitable industries in the country today. It is no wonder then that the DOHMH has an outstanding number of

high-paying positions, largely beating out runner-up, unionized agencies like the Department of Sanitation and the Department of Correction. Retaining talent in the public sector obviously requires providing competitive wages.

Question 3: Which jobs are the hardest to fill? Based on what?

Filling an open position can be a factor of many variables, including the firm's reputation or the office's distance from transportation. In the case of our data though the posting date is the most obvious indicator of which jobs are hardest to fill. Jobs that were posted farther in the past have probably been harder to fill.

To explore that possibility we'll create another variable called ConJobsDate, which will contain our data frame ordered from earliest to latest date.

```
ConJobsDate = arrange(ConJobs, as.Date(ConJobs$Posting.Date, format="%m
m/%dd/%YYYY %h:%m:%s"))

head(ConJobsDate)
```

```
##   Level                       Agency Salary.Range.From Salary.Range.
To
## 1    00 DEPARTMENT OF TRANSPORTATION               251              2
92
## 2    00 DEPARTMENT OF TRANSPORTATION               251              2
92
## 3    00 DEPARTMENT OF TRANSPORTATION               343              3
43
## 4    00 DEPARTMENT OF TRANSPORTATION               343              3
43
## 5    01     DEPARTMENT OF CORRECTION               356              3
92
## 6    01     DEPARTMENT OF CORRECTION               356              3
92
##   Salary.Frequency        Posting.Date
## 1            Daily 06/25/2014 00:00:00
## 2            Daily 06/25/2014 00:00:00
## 3            Daily 06/02/2015 00:00:00
## 4            Daily 06/02/2015 00:00:00
## 5            Daily 09/17/2015 00:00:00
## 6            Daily 09/17/2015 00:00:00
##                               Hours.Shift
## 1 40 hours / Monday-Friday; plus overtime
## 2 40 hours / Monday-Friday; plus overtime
## 3
## 4
## 5
## 6
##                         Work.Location.1 AvgSal
## 1 1 Bay Street, Staten Island, NY 10301  70590
## 2 1 Bay Street, Staten Island, NY 10301  70590
```

```
## 3                   88-26 Pitkin Avenue  89180
## 4                   88-26 Pitkin Avenue  89180
## 5                                        97240
## 6                                        97240
```

Let's pull the first 166 entries as a sample and get a count. These entries include jobs posted on or before November 2015, or about 4 months ago. According to authorities on public sector employment, it can take as much as four months to fill a government job.

The hypothesis is that the time it takes to fill a job is at least partially a function of the job's level, with higher level jobs that require more specialized talent taking longer to fill. Certainly other variables, like pay, could affect job fillings. But many of those variables may act in concordance with the job's level designation.

```
count(ConJobsDate[1:166,"Level"])
```

```
##    x freq
## 1 00   61
## 2 01   63
## 3 02   32
## 4 03   10
```

Let's compare that result to the whole population of jobs.

```
count(ConJobsDate["Level"])
```

```
##    Level freq
## 1     00 1172
## 2     01  679
## 3     02  757
## 4     03  358
## 5     04  171
## 6     1A    3
## 7     1B    2
## 8     3A    1
## 9     4A   32
## 10    4B   14
## 11    M1  410
## 12    M2  181
## 13    M3   85
## 14    M4   52
## 15    M5   24
## 16    M6    6
## 17    M7   14
## 18    M8    2
```

It would seem that lower level jobs, those between level 00 and 02, take the longest time to fill.

Further analysis should be conducted upon what variables influence the time it takes to fill a NYC Government position. It's possible that low salaries provide less of an incentive to take government jobs often located in lower Manhattan (and therefore further away from potential employees). It's also possible that there is a skills gap between those who would usually apply for these kinds of jobs and emerging job demands in the 21st century.