



---

## exploratory data analysis

---

### What is Exploratory Data Analysis?

**Exploratory data analysis (EDA)** is a technique used by data scientists to inspect, characterize and briefly summarize the contents of a dataset. EDA is often the first step when encountering a new or unfamiliar dataset. EDA helps the data scientist become acquainted with a dataset and test some basic assumptions about the data. By the end of the EDA process, some initial insights can be drawn from the dataset and a framework for further analysis or modeling is established.

---

## Civil Service List (Active)

**Dataset Analyzed:** *Civil Service List (Active)*

**About This Dataset:** A Civil Service List consists of all candidates who passed an exam, ranked in score order. An established list is considered active for no less than one year and no more than four years from the date of establishment. For more information visit DCAS' "Work for the City" webpage at: <https://www1.nyc.gov/site/dcas/employment/take-an-exam.page> Data provided by the Department of Citywide Administrative Services (DCAS), the City of New York: <https://data.cityofnewyork.us/City-Government/Civil-Service-List-Active-/vx8i-nprf>

**Acknowledgements:** NYC Open Data <https://opendata.cityofnewyork.us/>

**EDA Catalogue Number:** INS-010

**EDA Publication Date:** Tuesday, January 10, 2023

**Language:** Python

**Libraries Used:** NumPy, pandas, matplotlib, seaborn

**EDA Author:** David White

**Contact:** david@msmdesign.nyc | [msmdesign.nyc](http://msmdesign.nyc)

---

## 0. Prepare the workspace

### 0.1 Import Python libraries, packages and functions

```
In [2]: # import libraries for data wrangling, aggregate functions and basic descriptive statistics
import numpy as np
import pandas as pd

# import data visualization packages
import matplotlib.pyplot as plt
import seaborn as sns
```

### 0.2 Adjust display options to make plots easier to read and understand

```
In [4]: # specify seaborn styling options
sns.set_theme(
    context='talk',
    style='whitegrid',
    palette='Accent',
    font='Courier New',
    font_scale=1.15)

# allow plots to display inline within the notebook
%matplotlib inline
```

### 0.3 Set Markdown tables to align-left within notebook cells

```
In [5]: %%html
<style>
table {float:left}
</style>
```

### 0.4 Display all rows of output by default

```
In [6]: pd.set_option('display.max_rows', None)

# to reset:
# pd.reset_option('display.max_rows')
```

## 0.5 Format large numbers and display floating point values to two decimal places

```
In [7]: pd.set_option('display.float_format', '{:,.2f}'.format)

# to reset:
# pd.reset_option('display.float_format')
```

## 0.6 Load the raw data file into the notebook and visually confirm that it has been read in as expected

```
In [15]: # Load the data from a csv file (stored locally) into a new DataFrame object

csv = r"F:\Creative Cloud Files\MSM Client 001 - Mister Shepherd Media LLC\MSM Design\
candidates_temp = pd.read_csv(csv, encoding='utf-8', low_memory=False)
```

```
In [16]: # glimpse the first three rows

candidates_temp.head(3)
```

```
Out[16]:
```

	Exam No	List No	First Name	MI	Last Name	Adj. FA	List Title Code	List Title Desc	Group No	List Agency Code	List Age
0	162	877.00	YAHAIRA	NaN	ALMONTE	78.00	10001	ADMINISTRATIVE ACCOUNTANT	0	0	0
1	162	878.00	CHI SUN	NaN	CHOW	77.00	10001	ADMINISTRATIVE ACCOUNTANT	0	0	0
2	162	879.00	RACHEL	NaN	CANCEL	77.00	10001	ADMINISTRATIVE ACCOUNTANT	0	0	0

```
In [17]: # glimpse the last three rows

candidates_temp.tail(3)
```

```
Out[17]:
```

	Exam No	List No	First Name	MI	Last Name	Adj. FA	List Title Code	List Title Desc	Group No	List Agency Code	List Agen
568423	504	1.00	ANAND	NaN	HEMLALL	84.06	92610	MACHINIST	0	40	DEPARTM EDU
568424	504	1.00	MIHAIL	A	ZAINEA	78.19	92610	MACHINIST	0	826	DEPARTM ENVIRON PROT
568425	504	1.00	DIOGENES	A	MATA	73.10	92610	MACHINIST	0	841	DEPARTM TRANSPOF

```
In [18]: # glimpse ten randomly selected rows
candidates_temp.sample(10, random_state=42)
```

```
Out[18]:
```

	Exam No	List No	First Name	MI	Last Name	Adj. FA	List Title Code	List Title Desc	Group No
<b>386168</b>	6325	853.00	SHAKEEM	NaN	DAWKINS	72.50	70410	CORRECTION OFFICER	0
<b>339198</b>	7001	731.00	JAMES	R	MCGEE	104.00	70310	FIREFIGHTER	0
<b>501893</b>	9618	2,017.00	EDWIN	NaN	PATINO	81.67	91203	BUS OPERATOR	0
<b>80082</b>	9036	100.00	ISAAC	NaN	KHAN	75.00	20202	CIVIL ENGINEERING INTERN	0
<b>105489</b>	1177	309.00	RALEIGH	NaN	FOWLER	70.00	51001	SPECIAL CONSULTANT (MENTAL HEALTH STANDARDS AN...	0
<b>360019</b>	7001	10,593.00	JOSHUA	NaN	LORCY	97.00	70310	FIREFIGHTER	0
<b>16538</b>	9011	132.00	RATNAGIRISH	NaN	NORI	100.00	10050	COMPUTER SYSTEMS MANAGER	0
<b>171295</b>	9326	259.00	TIYANNA	NaN	CAMERON	88.23	60817	SCHOOL SAFETY AGENT	0
<b>293817</b>	5329	852.00	CEYHUN	NaN	DAGDEVIREN	84.71	70210	POLICE OFFICER	0
<b>431514</b>	144	2,242.00	INDIRA	A	SIMPSON	84.00	71012	POLICE COMMUNICATIONS TECHNICIAN	0

The data has been loaded and has been read in as expected.

## 0.7. Check the data type of each column

```
In [24]: # display the data type of each column
candidates_temp.dtypes
```

```
Out[24]: Exam No          int64
List No          float64
First Name      object
MI              object
Last Name       object
Adj. FA         float64
List Title Code int64
List Title Desc object
Group No        int64
List Agency Code int64
List Agency Desc object
List Div Code   float64
Published Date  object
Established Date object
Anniversary Date object
Extension Date  object
Veteran Credit  object
Parent Lgy Credit object
Sibling Lgy Credit object
Residency Credit object
dtype: object
```

**Most of the columns have been read in as *object* data type. We'll need to change the data type of some columns to a something more appropriate.**

**0.8 Refer to the [data dictionary](#) and make sure that our DataFrame's data types match the source data. Reassign data types where needed.**

```
In [25]: # cast column(s) containing dates to datetime data type

candidates_temp['Published Date'] = pd.to_datetime(candidates_temp['Published Date'])
candidates_temp['Established Date'] = pd.to_datetime(candidates_temp['Established Date'])
candidates_temp['Anniversary Date'] = pd.to_datetime(candidates_temp['Anniversary Date'])
candidates_temp['Extension Date'] = pd.to_datetime(candidates_temp['Extension Date'])
```

```
In [37]: # cast column(s) containing categorical variables to categorical data type

candidates_temp['Exam No'] = candidates_temp['Exam No'].astype('category')
candidates_temp['List Title Desc'] = candidates_temp['List Title Desc'].astype('category')
candidates_temp['List Agency Desc'] = candidates_temp['List Agency Desc'].astype('category')
candidates_temp['Veteran Credit'] = candidates_temp['Veteran Credit'].astype('category')
candidates_temp['Parent Lgy Credit'] = candidates_temp['Parent Lgy Credit'].astype('category')
candidates_temp['Sibling Lgy Credit'] = candidates_temp['Sibling Lgy Credit'].astype('category')
candidates_temp['Residency Credit'] = candidates_temp['Residency Credit'].astype('category')
```

```
In [38]: # display the DataFrame info once again to confirm that the data type changes have been made

candidates_temp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 568426 entries, 0 to 568425
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Exam No                568426 non-null  category
1   List No                568426 non-null  float64
2   First Name            568416 non-null  object
3   MI                    322663 non-null  object
4   Last Name             568418 non-null  object
5   Adj. FA               568426 non-null  float64
6   List Title Code       568426 non-null  int64
7   List Title Desc       568426 non-null  category
8   Group No              568426 non-null  int64
9   List Agency Code      568426 non-null  int64
10  List Agency Desc      568426 non-null  category
11  List Div Code         0 non-null       float64
12  Published Date        274960 non-null  datetime64[ns]
13  Established Date      557530 non-null  datetime64[ns]
14  Anniversary Date     557530 non-null  datetime64[ns]
15  Extension Date       258568 non-null  datetime64[ns]
16  Veteran Credit       13077 non-null   category
17  Parent Lgy Credit    1107 non-null    category
18  Sibling Lgy Credit   436 non-null     category
19  Residency Credit     119790 non-null  category
dtypes: category(7), datetime64[ns](4), float64(3), int64(3), object(3)
memory usage: 61.3+ MB
```

---

# 1. Describe the characteristics of the dataset

## 1.1 How many rows and how many columns are in our dataset?

```
In [28]: # display the number of rows and columns in the DataFrame

rows = candidates_temp.shape[0]
columns = candidates_temp.shape[1]

print(f'There are {rows} rows and {columns} columns in the dataset.')
```

There are 568426 rows and 20 columns in the dataset.

## 1.2 Identify the index of our DataFrame

```
In [29]: # display the index of the DataFrame

candidates_temp.index
```

```
Out[29]: RangeIndex(start=0, stop=568426, step=1)
```

Our DataFrame has an interger index. We know from the data dictionary that each row is an individual job candidate.

### 1.3 What are the column headings in our dataset?

```
In [30]: # display a List of the DataFrame's columns
```

```
list(candidates_temp.columns)
```

```
Out[30]: ['Exam No',  
         'List No',  
         'First Name',  
         'MI',  
         'Last Name',  
         'Adj. FA',  
         'List Title Code',  
         'List Title Desc',  
         'Group No',  
         'List Agency Code',  
         'List Agency Desc',  
         'List Div Code',  
         'Published Date',  
         'Established Date',  
         'Anniversary Date',  
         'Extension Date',  
         'Veteran Credit',  
         'Parent Lgy Credit',  
         'Sibling Lgy Credit',  
         'Residency Credit']
```

### 1.4 What are the data types of each column?

```
In [31]: # display the data type of each column in the DataFrame
```

```
candidates_temp.dtypes
```

```
Out[31]: Exam No                int64  
List No                float64  
First Name            object  
MI                    object  
Last Name             object  
Adj. FA               float64  
List Title Code       int64  
List Title Desc       category  
Group No              int64  
List Agency Code      int64  
List Agency Desc      category  
List Div Code         float64  
Published Date        datetime64[ns]  
Established Date      datetime64[ns]  
Anniversary Date      datetime64[ns]  
Extension Date        datetime64[ns]  
Veteran Credit        category  
Parent Lgy Credit     category  
Sibling Lgy Credit     category  
Residency Credit     category  
dtype: object
```

### 1.5 How many null values are in each column?

```
In [32]: # display the number of missing values in each column of the DataFrame
```

```
candidates_temp.isna().sum()
```

```
Out[32]: Exam No          0
List No           0
First Name       10
MI              245763
Last Name        8
Adj. FA          0
List Title Code  0
List Title Desc  0
Group No         0
List Agency Code 0
List Agency Desc 0
List Div Code    568426
Published Date   293466
Established Date 10896
Anniversary Date 10896
Extension Date   309858
Veteran Credit   555349
Parent Lgy Credit 567319
Sibling Lgy Credit 567990
Residency Credit 448636
dtype: int64
```

## 1.6 How many unique values are there in each column?

```
In [33]: # display the count of unique elements in each column
```

```
candidates_temp.nunique(axis=0, dropna=True)
```

```
Out[33]: Exam No          861
List No           75403
First Name       58600
MI               37
Last Name        83436
Adj. FA          2869
List Title Code  490
List Title Desc  490
Group No         3
List Agency Code 62
List Agency Desc 63
List Div Code    0
Published Date   165
Established Date 311
Anniversary Date 315
Extension Date   116
Veteran Credit   2
Parent Lgy Credit 1
Sibling Lgy Credit 1
Residency Credit 1
dtype: int64
```

---

## 2. Briefly summarize the contents of the dataset



## 2.1 Summarize the columns containing numerical variables

```
In [34]: # describe numeric columns only

num_cols = ['Adj. FA']

candidates_temp[num_cols].describe(include=[np.number])
```

```
Out[34]:
```

	Adj. FA
<b>count</b>	568,426.00
<b>mean</b>	87.52
<b>std</b>	8.91
<b>min</b>	65.00
<b>25%</b>	80.75
<b>50%</b>	88.23
<b>75%</b>	94.75
<b>max</b>	127.24

## 2.2 Summarize the columns containing datetime variables

```
In [35]: # summarize the data contained in columns with the 'datetime' data type only

date_cols = ['Published Date',
             'Established Date',
             'Anniversary Date',
             'Extension Date']

candidates_temp[date_cols].describe(datetime_is_numeric=True)
```

```
Out[35]:
```

	Published Date	Established Date	Anniversary Date	Extension Date
<b>count</b>	274960	557530	557530	258568
<b>mean</b>	2017-11-26 11:34:59.563573248	2019-02-07 12:36:10.344197120	2023-02-07 12:55:35.712876800	2023-06-19 07:19:32.636985600
<b>min</b>	2010-07-28 00:00:00	2013-10-30 00:00:00	2017-10-30 00:00:00	2023-01-11 00:00:00
<b>25%</b>	2015-09-23 00:00:00	2017-04-05 00:00:00	2021-04-05 00:00:00	2023-03-16 00:00:00
<b>50%</b>	2018-04-11 00:00:00	2019-02-27 00:00:00	2023-02-27 00:00:00	2023-04-25 00:00:00
<b>75%</b>	2019-02-27 00:00:00	2021-01-13 00:00:00	2025-01-13 00:00:00	2023-09-22 00:00:00
<b>max</b>	2022-11-23 00:00:00	2023-01-11 00:00:00	2027-01-11 00:00:00	2024-01-31 00:00:00

## 2.3 Summarize the columns containing categorical variables

```
In [39]: # summarize the data contained in columns with the 'category' data type only
```

```
candidates_temp.describe(include=['category'])
```

Out[39]:

	Exam No	List Title Desc	List Agency Desc	Veteran Credit	Parent Lgy Credit	Sibling Lgy Credit	Residency Credit
<b>count</b>	568426	568426	568426	13077	1107	436	119790
<b>unique</b>	861	490	63	2	1	1	1
<b>top</b>	5001	POLICE OFFICER	OPEN COMPETITIVE	Veteran's Credit	Parent Legacy Credit	Sibling Legacy Credit	Residency Credit
<b>freq</b>	68859	80635	525625	11370	1107	436	119790

## 3. Select a subset of data for closer examination

### 3.1 Select a subset of columns

```
In [ ]: # display all columns
```

```
list(df.columns)
```

```
In [84]: # select a subset of columns to examine
```

```
selected_cols = ['Exam No',  
                'List Title Desc',  
                'List Agency Desc',  
                'Published Date',  
                'Established Date',  
                'Anniversary Date',  
                'Extension Date',  
                'Veteran Credit',  
                'Parent Lgy Credit',  
                'Sibling Lgy Credit',  
                'Residency Credit',  
                'Adj. FA']
```

```
candidates = candidates_temp[selected_cols]
```

### 3.2 Select a subset of rows

```
In [85]: # subset only exams results published in the 2022 calendar year
```

```
candidates2022 = candidates.loc[candidates['Published Date'] >= '2022-01-01 00:00:00']
```

```
In [86]: # confirm that the date range has been subset as expected
```

```
candidates2022['Published Date'].min()
```

```
Out[86]: Timestamp('2022-02-02 00:00:00')
```

```
In [87]: candidates2022['Published Date'].max()
```

```
Out[87]: Timestamp('2022-11-23 00:00:00')
```

### 3.3 Display the shape of the data subset

```
In [89]: rows = candidates2022.shape[0]
columns = candidates2022.shape[1]

print(f'There are {rows} rows and {columns} columns in the subset.')
```

There are 24972 rows and 12 columns in the subset.

---

## 4. Examine the individual variables in the dataset

### 4.1 What is the distribution of job titles represented in the dataset?

```
In [68]: candidates2022['List Title Desc'].value_counts().nlargest(40)
```

```

Out[68]: ADMINISTRATIVE MANAGER 6349
PRINCIPAL ADMINISTRATIVE ASSOCIATE 4353
ADMINISTRATIVE STAFF ANALYST 2455
SCHOOL SAFETY AGENT 1445
COMPUTER ASSOCIATE (SOFTWARE) 759
FRAUD INVESTIGATOR 640
CERTIFIED IT ADMINISTRATOR (LAN/WAN) 640
COMPUTER ASSOCIATE (TECHNICAL SUPPORT) 624
YOUTH DEVELOPMENT SPECIALIST 594
CALL CENTER REPRESENTATIVE 594
CONSTRUCTION PROJECT MANAGER 587
EMERGENCY MEDICAL SPECIALIST-PARAMEDIC 568
QUALITY ASSURANCE SPECIALIST TRAINEE 461
ASSOCIATE RETIREMENT BENEFITS EXAMINER 458
COMPUTER PROGRAMMER ANALYST 419
CONTRACT SPECIALIST 356
ADMINISTRATIVE QUALITY ASSURANCE SPECIALIST 326
PUBLIC HEALTH SANITARIAN 317
INSPECTOR (HOUSING) 276
ADMINISTRATIVE HOUSING DEVELOPMENT SPECIALIST 256
CITY TAX AUDITOR 243
COMPUTER PROGRAMMER ANALYST TRAINEE 237
INSPECTOR (CONSTRUCTION) 229
CHILD AND FAMILY SPECIALIST 200
SPECIAL OFFICER 187
ASSOCIATE PARK SERVICE WORKER 178
ADMINISTRATIVE CITY PLANNER 177
SUPERVISOR OF SCHOOL SECURITY 166
HIGHWAYS AND SEWERS INSPECTOR 154
ELECTRICIAN 141
INDUSTRIAL HYGIENIST 137
ADMINISTRATIVE INSPECTOR (ELECTRICAL) 117
QUALITY ASSURANCE SPECIALIST (BUILDING REPAIRS) 100
WATER USE INSPECTOR 90
PLAN EXAMINER (BUILDINGS) 39
MECHANICAL ENGINEER 31
PHYSICAL THERAPIST (DOE) 22
INSPECTOR (HOISTS AND RIGGING) 19
MARINE ENGINEER (DIESEL) 17
HEATING PLANT TECHNICIAN (HOUSING AUTHORITY) 11
Name: List Title Desc, dtype: int64

```

```

In [69]: candidates2022['List Title Desc'].value_counts(normalize=True).nlargest(40)

```

```

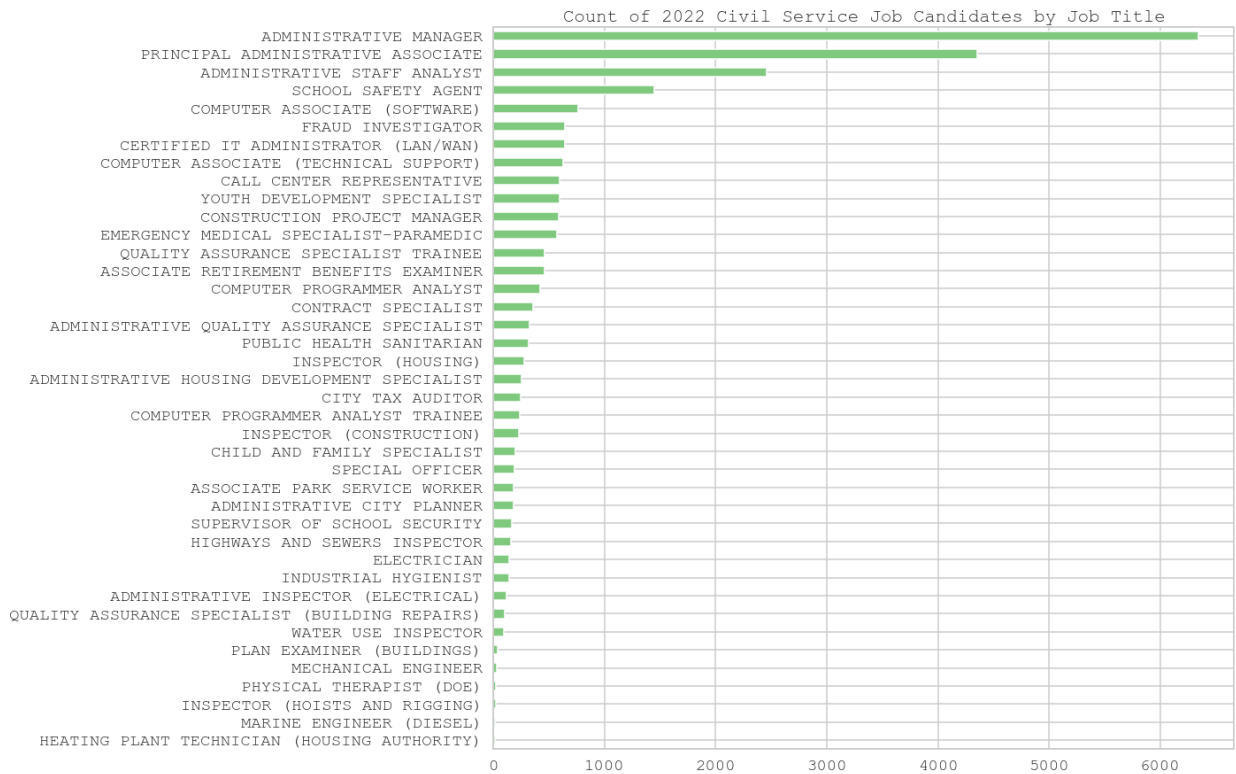
Out[69]: ADMINISTRATIVE MANAGER 0.25
PRINCIPAL ADMINISTRATIVE ASSOCIATE 0.17
ADMINISTRATIVE STAFF ANALYST 0.10
SCHOOL SAFETY AGENT 0.06
COMPUTER ASSOCIATE (SOFTWARE) 0.03
FRAUD INVESTIGATOR 0.03
CERTIFIED IT ADMINISTRATOR (LAN/WAN) 0.03
COMPUTER ASSOCIATE (TECHNICAL SUPPORT) 0.02
YOUTH DEVELOPMENT SPECIALIST 0.02
CALL CENTER REPRESENTATIVE 0.02
CONSTRUCTION PROJECT MANAGER 0.02
EMERGENCY MEDICAL SPECIALIST-PARAMEDIC 0.02
QUALITY ASSURANCE SPECIALIST TRAINEE 0.02
ASSOCIATE RETIREMENT BENEFITS EXAMINER 0.02
COMPUTER PROGRAMMER ANALYST 0.02
CONTRACT SPECIALIST 0.01
ADMINISTRATIVE QUALITY ASSURANCE SPECIALIST 0.01
PUBLIC HEALTH SANITARIAN 0.01
INSPECTOR (HOUSING) 0.01
ADMINISTRATIVE HOUSING DEVELOPMENT SPECIALIST 0.01
CITY TAX AUDITOR 0.01
COMPUTER PROGRAMMER ANALYST TRAINEE 0.01
INSPECTOR (CONSTRUCTION) 0.01
CHILD AND FAMILY SPECIALIST 0.01
SPECIAL OFFICER 0.01
ASSOCIATE PARK SERVICE WORKER 0.01
ADMINISTRATIVE CITY PLANNER 0.01
SUPERVISOR OF SCHOOL SECURITY 0.01
HIGHWAYS AND SEWERS INSPECTOR 0.01
ELECTRICIAN 0.01
INDUSTRIAL HYGIENIST 0.01
ADMINISTRATIVE INSPECTOR (ELECTRICAL) 0.00
QUALITY ASSURANCE SPECIALIST (BUILDING REPAIRS) 0.00
WATER USE INSPECTOR 0.00
PLAN EXAMINER (BUILDINGS) 0.00
MECHANICAL ENGINEER 0.00
PHYSICAL THERAPIST (DOE) 0.00
INSPECTOR (HOISTS AND RIGGING) 0.00
MARINE ENGINEER (DIESEL) 0.00
HEATING PLANT TECHNICIAN (HOUSING AUTHORITY) 0.00
Name: List Title Desc, dtype: float64

```

```

In [76]: candidates2022['List Title Desc'].value_counts().nlargest(40).sort_values().plot(kind=
figsi
title

```



In [ ]:

## 4.2 What is the distribution of hiring agencies represented in the dataset?

In [99]:

```
candidates2022['List Agency Desc'].value_counts().dropna()
```

Out[99]:	OPEN COMPETITIVE	22236
	FIRE DEPARTMENT	641
	HRA/DEPARTMENT OF SOCIAL SERVICES	343
	POLICE DEPARTMENT	282
	DEPARTMENT OF PARKS & RECREATION	261
	DEPARTMENT OF ENVIRONMENTAL PROTECTION	158
	ADMINISTRATION FOR CHILDREN'S SERVICES	124
	DEPARTMENT OF HEALTH AND MENTAL HYGIENE	115
	DEPARTMENT OF FINANCE	114
	DEPARTMENT OF TRANSPORTATION	105
	DEPARTMENT OF EDUCATION	61
	DEPARTMENT OF INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS	60
	DEPARTMENT OF CITYWIDE ADMINISTRATIVE SERVICES	57
	HOUSING PRESERVATION & DEVELOPMENT	49
	DEPARTMENT OF CORRECTION	48
	N.Y.C. HOUSING AUTHORITY	40
	DEPARTMENT OF BUILDINGS	36
	DEPARTMENT OF SANITATION	32
	DEPARTMENT OF HOMELESS SERVICES	19
	DEPARTMENT OF DESIGN AND CONSTRUCTION	17
	OFFICE OF PAYROLL ADMINISTRATION	15
	N.Y.C. TRANSIT AUTHORITY	13
	NYC EMPLOYEES' RETIREMENT SYSTEM	12
	TAXI AND LIMOUSINE COMMISSION	12
	OFFICE OF THE COMPTROLLER	12
	DEPARTMENT OF PROBATION	10
	DEPARTMENT OF YOUTH AND COMMUNITY DEVELOPMENT	10
	DISTRICT ATTORNEY-KINGS COUNTY	9
	FINANCIAL INFORMATION SERVICES AGENCY	8
	OFFICE OF ADMINISTRATIVE TRIALS AND HEARINGS (OATH)	7
	DEPARTMENT OF CONSUMER AFFAIRS	6
	DEPARTMENT FOR THE AGING	6
	TEACHERS' RETIREMENT SYSTEM	6
	NYC POLICE PENSION FUND	5
	DISTRICT ATTORNEY-BRONX COUNTY	5
	DISTRICT ATTORNEY-QUEENS COUNTY	5
	DEPARTMENT OF SMALL BUSINESS SERVICES	4
	DEPARTMENT OF INVESTIGATION	3
	OFFICE OF LABOR RELATIONS	3
	DEPARTMENT OF CITY PLANNING	2
	TAX COMMISSION	2
	BOROUGH PRESIDENT-BROOKLYN	2
	DEPT. OF RECORDS AND INFORMATION SERVICES	2
	LAW DEPARTMENT	2
	OFFICE OF MANAGEMENT AND BUDGET	2
	NEW YORK CITY FIRE PENSION FUND	2
	COMMUNITY BOARD NO.3-BRONX	1
	CITY CLERK	1
	CIVIL SERVICE COMMISSION	1
	CIVILIAN COMPLAINT REVIEW BOARD	1
	DISTRICT ATTORNEY-RICHMOND COUNTY	1
	DEPARTMENT OF VETERANS' SERVICES	1
	INDEPENDENT BUDGET OFFICE	1
	DISTRICT ATTORNEY-NEW YORK COUNTY	1
	TRIBOROUGH BRIDGE AND TUNNEL AUTHORITY	1
	EQUAL EMPLOYMENT PRACTICES COMMISSION	0
	COMMISSION ON HUMAN RIGHTS	0
	OFFICE OF THE ACTUARY	0
	MAYOR'S OFFICE OF CONTRACT SERVICES	0
	LANDMARKS PRESERVATION COMMISSION	0

QUALIFIED INCUMBENT	0
BOROUGH PRESIDENT-QUEENS	0
BOROUGH PRESIDENT-MANHATTAN	0

Name: List Agency Desc, dtype: int64

```
In [98]: candidates2022['List Agency Desc'].value_counts(normalize=True).dropna()
```



Out[98]:	OPEN COMPETITIVE	0.89
	FIRE DEPARTMENT	0.03
	HRA/DEPARTMENT OF SOCIAL SERVICES	0.01
	POLICE DEPARTMENT	0.01
	DEPARTMENT OF PARKS & RECREATION	0.01
	DEPARTMENT OF ENVIRONMENTAL PROTECTION	0.01
	ADMINISTRATION FOR CHILDREN'S SERVICES	0.00
	DEPARTMENT OF HEALTH AND MENTAL HYGIENE	0.00
	DEPARTMENT OF FINANCE	0.00
	DEPARTMENT OF TRANSPORTATION	0.00
	DEPARTMENT OF EDUCATION	0.00
	DEPARTMENT OF INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS	0.00
	DEPARTMENT OF CITYWIDE ADMINISTRATIVE SERVICES	0.00
	HOUSING PRESERVATION & DEVELOPMENT	0.00
	DEPARTMENT OF CORRECTION	0.00
	N.Y.C. HOUSING AUTHORITY	0.00
	DEPARTMENT OF BUILDINGS	0.00
	DEPARTMENT OF SANITATION	0.00
	DEPARTMENT OF HOMELESS SERVICES	0.00
	DEPARTMENT OF DESIGN AND CONSTRUCTION	0.00
	OFFICE OF PAYROLL ADMINISTRATION	0.00
	N.Y.C. TRANSIT AUTHORITY	0.00
	NYC EMPLOYEES' RETIREMENT SYSTEM	0.00
	TAXI AND LIMOUSINE COMMISSION	0.00
	OFFICE OF THE COMPTROLLER	0.00
	DEPARTMENT OF PROBATION	0.00
	DEPARTMENT OF YOUTH AND COMMUNITY DEVELOPMENT	0.00
	DISTRICT ATTORNEY-KINGS COUNTY	0.00
	FINANCIAL INFORMATION SERVICES AGENCY	0.00
	OFFICE OF ADMINISTRATIVE TRIALS AND HEARINGS (OATH)	0.00
	DEPARTMENT OF CONSUMER AFFAIRS	0.00
	DEPARTMENT FOR THE AGING	0.00
	TEACHERS' RETIREMENT SYSTEM	0.00
	NYC POLICE PENSION FUND	0.00
	DISTRICT ATTORNEY-BRONX COUNTY	0.00
	DISTRICT ATTORNEY-QUEENS COUNTY	0.00
	DEPARTMENT OF SMALL BUSINESS SERVICES	0.00
	DEPARTMENT OF INVESTIGATION	0.00
	OFFICE OF LABOR RELATIONS	0.00
	DEPARTMENT OF CITY PLANNING	0.00
	TAX COMMISSION	0.00
	BOROUGH PRESIDENT-BROOKLYN	0.00
	DEPT. OF RECORDS AND INFORMATION SERVICES	0.00
	LAW DEPARTMENT	0.00
	OFFICE OF MANAGEMENT AND BUDGET	0.00
	NEW YORK CITY FIRE PENSION FUND	0.00
	COMMUNITY BOARD NO.3-BRONX	0.00
	CITY CLERK	0.00
	CIVIL SERVICE COMMISSION	0.00
	CIVILIAN COMPLAINT REVIEW BOARD	0.00
	DISTRICT ATTORNEY-RICHMOND COUNTY	0.00
	DEPARTMENT OF VETERANS' SERVICES	0.00
	INDEPENDENT BUDGET OFFICE	0.00
	DISTRICT ATTORNEY-NEW YORK COUNTY	0.00
	TRIBOROUGH BRIDGE AND TUNNEL AUTHORITY	0.00
	EQUAL EMPLOYMENT PRACTICES COMMISSION	0.00
	COMMISSION ON HUMAN RIGHTS	0.00
	OFFICE OF THE ACTUARY	0.00
	MAYOR'S OFFICE OF CONTRACT SERVICES	0.00
	LANDMARKS PRESERVATION COMMISSION	0.00

```

QUALIFIED INCUMBENT                0.00
BOROUGH PRESIDENT-QUEENS           0.00
BOROUGH PRESIDENT-MANHATTAN        0.00
Name: List Agency Desc, dtype: float64

```

In [102]...

```

candidates2022['List Agency Desc'].value_counts().dropna().sort_values().plot(kind='bar',
figsi
title

```



### 4.3 What is the distribution of exam scores represented in the dataset?

In [91]:

```

# display summary statistics for exam scores (overall)
candidates2022['Adj. FA'].describe()

```

```
Out[91]: count    24,972.00
         mean      91.20
         std       10.84
         min       70.00
         25%      81.82
         50%     100.00
         75%     100.00
         max      120.00
         Name: Adj. FA, dtype: float64
```

```
In [97]: # display summary statistics for exam scores by job title
         candidates2022.groupby('List Title Desc').describe().dropna()
```

Out[97]:

	count	mean	std	min	25%	50%	75%	max	Adj. FA
<b>ADMINISTRATIVE CITY PLANNER</b>	177.00	99.05	3.79	70.00	100.00	100.00	100.00	100.00	
<b>ADMINISTRATIVE HOUSING DEVELOPMENT SPECIALIST</b>	256.00	93.75	10.32	70.00	90.00	100.00	100.00	100.00	
<b>ADMINISTRATIVE INSPECTOR (ELECTRICAL)</b>	117.00	96.91	6.63	70.00	100.00	100.00	100.00	105.00	
<b>ADMINISTRATIVE MANAGER</b>	6,349.00	97.95	5.62	70.00	100.00	100.00	100.00	120.00	
<b>ADMINISTRATIVE QUALITY ASSURANCE SPECIALIST</b>	326.00	97.56	6.94	70.00	100.00	100.00	100.00	110.00	
<b>ADMINISTRATIVE STAFF ANALYST</b>	2,455.00	78.96	5.49	70.00	74.38	78.12	83.12	97.50	
<b>ASSOCIATE PARK SERVICE WORKER</b>	178.00	97.71	4.90	77.50	97.00	100.00	100.00	100.00	
<b>ASSOCIATE RETIREMENT BENEFITS EXAMINER</b>	458.00	87.97	10.29	70.00	80.00	86.25	100.00	100.00	
<b>CALL CENTER REPRESENTATIVE</b>	594.00	91.12	10.82	70.00	85.00	97.00	100.00	120.00	
<b>CERTIFIED IT ADMINISTRATOR (LAN/WAN)</b>	640.00	94.68	9.49	70.00	94.00	100.00	100.00	120.00	
<b>CHILD AND FAMILY SPECIALIST</b>	200.00	86.19	9.93	70.00	77.67	85.00	92.50	110.67	
<b>CITY TAX AUDITOR</b>	243.00	76.52	10.02	70.00	70.00	70.00	80.00	100.00	
<b>COMPUTER ASSOCIATE (SOFTWARE)</b>	759.00	93.46	9.03	70.00	87.12	100.00	100.00	105.00	
<b>COMPUTER ASSOCIATE (TECHNICAL SUPPORT)</b>	624.00	94.77	8.09	70.00	90.00	100.00	100.00	110.00	
<b>COMPUTER PROGRAMMER ANALYST</b>	419.00	89.70	11.23	70.00	78.88	94.25	100.00	110.00	
<b>COMPUTER PROGRAMMER ANALYST TRAINEE</b>	237.00	83.95	6.63	70.01	79.68	84.52	88.39	98.23	
<b>CONSTRUCTION PROJECT MANAGER</b>	587.00	78.36	5.88	70.00	74.00	78.00	82.00	99.00	
<b>CONTRACT SPECIALIST</b>	356.00	93.28	9.25	72.50	85.00	100.00	100.00	110.00	
<b>ELECTRICIAN</b>	141.00	80.50	6.98	70.00	73.75	81.25	85.00	97.50	
<b>EMERGENCY MEDICAL SPECIALIST-PARAMEDIC</b>	568.00	85.91	9.87	70.00	77.00	86.00	95.00	100.00	
<b>FRAUD INVESTIGATOR</b>	640.00	81.16	12.61	70.00	70.00	75.00	95.00	115.00	
<b>HEATING PLANT TECHNICIAN (HOUSING AUTHORITY)</b>	11.00	79.45	6.76	70.00	74.00	76.00	85.00	90.00	
<b>HIGHWAYS AND SEWERS INSPECTOR</b>	154.00	95.83	9.29	70.00	100.00	100.00	100.00	110.00	
<b>INDUSTRIAL HYGIENIST</b>	137.00	87.31	11.50	70.00	75.83	87.50	100.00	100.00	
<b>INSPECTOR (CONSTRUCTION)</b>	229.00	95.88	8.47	70.00	98.00	100.00	100.00	105.00	
<b>INSPECTOR (HOISTS AND RIGGING)</b>	19.00	83.34	9.29	71.00	77.00	81.00	91.50	100.00	

	count	mean	std	min	25%	50%	75%	max
<b>INSPECTOR (HOUSING)</b>	276.00	92.68	11.38	70.00	85.74	100.00	100.00	105.00
<b>MARINE ENGINEER (DIESEL)</b>	17.00	84.38	14.39	70.00	70.00	75.00	100.00	100.00
<b>MECHANICAL ENGINEER</b>	31.00	87.90	11.51	70.00	76.00	91.00	100.00	100.00
<b>PHYSICAL THERAPIST (DOE)</b>	22.00	89.36	7.85	73.50	85.00	85.00	96.75	100.00
<b>PLAN EXAMINER (BUILDINGS)</b>	39.00	86.87	13.12	70.00	73.00	90.50	100.00	100.00
<b>PRINCIPAL ADMINISTRATIVE ASSOCIATE</b>	4,353.00	97.23	7.07	70.00	100.00	100.00	100.00	110.00
<b>PUBLIC HEALTH SANITARIAN</b>	317.00	78.61	10.58	70.00	70.00	73.12	85.00	100.00
<b>QUALITY ASSURANCE SPECIALIST (BUILDING REPAIRS)</b>	100.00	94.85	9.71	70.00	93.62	100.00	100.00	105.00
<b>QUALITY ASSURANCE SPECIALIST TRAINEE</b>	461.00	86.41	11.24	70.00	77.50	83.30	100.00	115.00
<b>SCHOOL SAFETY AGENT</b>	1,445.00	82.16	7.00	70.91	76.36	81.82	87.27	101.36
<b>SPECIAL OFFICER</b>	187.00	82.98	6.77	70.00	77.50	83.75	88.75	101.25
<b>SUPERVISOR OF SCHOOL SECURITY</b>	166.00	78.56	3.76	71.80	75.56	77.69	80.88	89.38
<b>WATER USE INSPECTOR</b>	90.00	86.08	11.99	70.00	73.50	86.25	100.00	105.00
<b>YOUTH DEVELOPMENT SPECIALIST</b>	594.00	86.36	12.91	70.00	70.00	88.38	100.00	110.00

In [105...

```
# display summary statistics for exam scores by hiring department
candidates2022.groupby('List Agency Desc').describe().dropna()
```

Out[105]:

	count	mean	std	min	25%	50%	75%	max
<b>ADMINISTRATION FOR CHILDREN'S SERVICES</b>	124.00	98.61	4.18	76.00	100.00	100.00	100.00	105.00
<b>BOROUGH PRESIDENT-BROOKLYN</b>	2.00	96.25	5.30	92.50	94.38	96.25	98.12	100.00
<b>DEPARTMENT FOR THE AGING</b>	6.00	99.25	1.84	95.50	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF BUILDINGS</b>	36.00	99.38	2.75	85.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF CITY PLANNING</b>	2.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF CITYWIDE ADMINISTRATIVE SERVICES</b>	57.00	98.14	5.09	74.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF CONSUMER AFFAIRS</b>	6.00	97.00	7.35	82.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF CORRECTION</b>	48.00	97.61	6.32	70.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF DESIGN AND CONSTRUCTION</b>	17.00	99.43	2.36	90.25	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF EDUCATION</b>	61.00	97.91	6.26	70.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF ENVIRONMENTAL PROTECTION</b>	158.00	99.28	3.22	73.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF FINANCE</b>	114.00	98.68	4.10	76.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF HEALTH AND MENTAL HYGIENE</b>	115.00	98.40	4.69	73.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF HOMELESS SERVICES</b>	19.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS</b>	60.00	98.65	4.87	70.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF INVESTIGATION</b>	3.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF PARKS &amp; RECREATION</b>	261.00	97.86	5.25	70.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF PROBATION</b>	10.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF SANITATION</b>	32.00	96.16	8.70	70.00	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF SMALL BUSINESS SERVICES</b>	4.00	98.75	2.50	95.00	98.75	100.00	100.00	100.00
<b>DEPARTMENT OF TRANSPORTATION</b>	105.00	98.59	4.08	77.50	100.00	100.00	100.00	100.00
<b>DEPARTMENT OF YOUTH AND COMMUNITY DEVELOPMENT</b>	10.00	99.70	0.95	97.00	100.00	100.00	100.00	100.00
<b>DEPT. OF RECORDS AND INFORMATION SERVICES</b>	2.00	94.50	7.78	89.00	91.75	94.50	97.25	100.00

								Adj. FA
	count	mean	std	min	25%	50%	75%	max
List Agency Desc								
DISTRICT ATTORNEY-BRONX COUNTY	5.00	99.60	0.89	98.00	100.00	100.00	100.00	100.00
DISTRICT ATTORNEY-KINGS COUNTY	9.00	98.00	6.00	82.00	100.00	100.00	100.00	100.00
DISTRICT ATTORNEY-QUEENS COUNTY	5.00	97.60	5.37	88.00	100.00	100.00	100.00	100.00
FINANCIAL INFORMATION SERVICES AGENCY	8.00	98.38	4.60	87.00	100.00	100.00	100.00	100.00
FIRE DEPARTMENT	641.00	87.50	10.29	70.00	78.00	88.00	100.00	100.00
HOUSING PRESERVATION & DEVELOPMENT	49.00	97.38	7.18	70.00	100.00	100.00	100.00	100.00
HRA/DEPARTMENT OF SOCIAL SERVICES	343.00	97.53	6.23	70.00	100.00	100.00	100.00	105.00
LAW DEPARTMENT	2.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00
N.Y.C. HOUSING AUTHORITY	40.00	97.53	4.93	82.38	99.25	100.00	100.00	100.00
N.Y.C. TRANSIT AUTHORITY	13.00	99.65	0.90	97.00	100.00	100.00	100.00	100.00
NEW YORK CITY FIRE PENSION FUND	2.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00
NYC EMPLOYEES' RETIREMENT SYSTEM	12.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00
NYC POLICE PENSION FUND	5.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00
OFFICE OF ADMINISTRATIVE TRIALS AND HEARINGS (OATH)	7.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00
OFFICE OF LABOR RELATIONS	3.00	90.00	17.32	70.00	85.00	100.00	100.00	100.00
OFFICE OF MANAGEMENT AND BUDGET	2.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00
OFFICE OF PAYROLL ADMINISTRATION	15.00	96.58	7.08	80.00	98.50	100.00	100.00	100.00
OFFICE OF THE COMPTROLLER	12.00	98.79	4.19	85.50	100.00	100.00	100.00	100.00
OPEN COMPETITIVE	22,236.00	90.79	10.98	70.00	81.00	98.50	100.00	120.00
POLICE DEPARTMENT	282.00	87.07	10.78	71.80	76.62	82.66	100.00	100.00
TAX COMMISSION	2.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00
TAXI AND LIMOUSINE COMMISSION	12.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00
TEACHERS' RETIREMENT SYSTEM	6.00	96.00	7.27	82.00	95.50	100.00	100.00	100.00

In [133]...

```
# find the job titles with the top mean scores
```

```
candidates2022.groupby('List Title Desc')['Adj. FA'].mean().dropna().sort_values(ascr
```

Out[133]:

List Title Desc	
ADMINISTRATIVE CITY PLANNER	99.05
ADMINISTRATIVE MANAGER	97.95
ASSOCIATE PARK SERVICE WORKER	97.71
ADMINISTRATIVE QUALITY ASSURANCE SPECIALIST	97.56
PRINCIPAL ADMINISTRATIVE ASSOCIATE	97.23
ADMINISTRATIVE INSPECTOR (ELECTRICAL)	96.91
INSPECTOR (CONSTRUCTION)	95.88
HIGHWAYS AND SEWERS INSPECTOR	95.83
QUALITY ASSURANCE SPECIALIST (BUILDING REPAIRS)	94.85
COMPUTER ASSOCIATE (TECHNICAL SUPPORT)	94.77
CERTIFIED IT ADMINISTRATOR (LAN/WAN)	94.68
ADMINISTRATIVE HOUSING DEVELOPMENT SPECIALIST	93.75
COMPUTER ASSOCIATE (SOFTWARE)	93.46
CONTRACT SPECIALIST	93.28
INSPECTOR (HOUSING)	92.68
CALL CENTER REPRESENTATIVE	91.12
COMPUTER PROGRAMMER ANALYST	89.70
PHYSICAL THERAPIST (DOE)	89.36
ASSOCIATE RETIREMENT BENEFITS EXAMINER	87.97
MECHANICAL ENGINEER	87.90
INDUSTRIAL HYGIENIST	87.31
PLAN EXAMINER (BUILDINGS)	86.87
QUALITY ASSURANCE SPECIALIST TRAINEE	86.41
YOUTH DEVELOPMENT SPECIALIST	86.36
CHILD AND FAMILY SPECIALIST	86.19
WATER USE INSPECTOR	86.08
EMERGENCY MEDICAL SPECIALIST-PARAMEDIC	85.91
MARINE ENGINEER (DIESEL)	84.38
COMPUTER PROGRAMMER ANALYST TRAINEE	83.95
INSPECTOR (HOISTS AND RIGGING)	83.34
SPECIAL OFFICER	82.98
SCHOOL SAFETY AGENT	82.16
FRAUD INVESTIGATOR	81.16
ELECTRICIAN	80.50
HEATING PLANT TECHNICIAN (HOUSING AUTHORITY)	79.45
ADMINISTRATIVE STAFF ANALYST	78.96
PUBLIC HEALTH SANITARIAN	78.61
SUPERVISOR OF SCHOOL SECURITY	78.56
CONSTRUCTION PROJECT MANAGER	78.36
CITY TAX AUDITOR	76.52

Name: Adj. FA, dtype: float64

In [130]...

```
# find the job titles with the top median scores
```

```
candidates2022.groupby('List Title Desc')['Adj. FA'].median().dropna().sort_values(ascr
```



```

Out[130]: List Title Desc
ADMINISTRATIVE CITY PLANNER 100.00
ADMINISTRATIVE HOUSING DEVELOPMENT SPECIALIST 100.00
CONTRACT SPECIALIST 100.00
INSPECTOR (CONSTRUCTION) 100.00
INSPECTOR (HOUSING) 100.00
COMPUTER ASSOCIATE (TECHNICAL SUPPORT) 100.00
COMPUTER ASSOCIATE (SOFTWARE) 100.00
CERTIFIED IT ADMINISTRATOR (LAN/WAN) 100.00
PRINCIPAL ADMINISTRATIVE ASSOCIATE 100.00
ASSOCIATE PARK SERVICE WORKER 100.00
QUALITY ASSURANCE SPECIALIST (BUILDING REPAIRS) 100.00
ADMINISTRATIVE QUALITY ASSURANCE SPECIALIST 100.00
ADMINISTRATIVE MANAGER 100.00
ADMINISTRATIVE INSPECTOR (ELECTRICAL) 100.00
HIGHWAYS AND SEWERS INSPECTOR 100.00
CALL CENTER REPRESENTATIVE 97.00
COMPUTER PROGRAMMER ANALYST 94.25
MECHANICAL ENGINEER 91.00
PLAN EXAMINER (BUILDINGS) 90.50
YOUTH DEVELOPMENT SPECIALIST 88.38
INDUSTRIAL HYGIENIST 87.50
WATER USE INSPECTOR 86.25
ASSOCIATE RETIREMENT BENEFITS EXAMINER 86.25
EMERGENCY MEDICAL SPECIALIST-PARAMEDIC 86.00
PHYSICAL THERAPIST (DOE) 85.00
CHILD AND FAMILY SPECIALIST 85.00
COMPUTER PROGRAMMER ANALYST TRAINEE 84.52
SPECIAL OFFICER 83.75
QUALITY ASSURANCE SPECIALIST TRAINEE 83.30
SCHOOL SAFETY AGENT 81.82
ELECTRICIAN 81.25
INSPECTOR (HOISTS AND RIGGING) 81.00
ADMINISTRATIVE STAFF ANALYST 78.12
CONSTRUCTION PROJECT MANAGER 78.00
SUPERVISOR OF SCHOOL SECURITY 77.69
HEATING PLANT TECHNICIAN (HOUSING AUTHORITY) 76.00
MARINE ENGINEER (DIESEL) 75.00
FRAUD INVESTIGATOR 75.00
PUBLIC HEALTH SANITARIAN 73.12
CITY TAX AUDITOR 70.00
Name: Adj. FA, dtype: float64

```

```

In [131... # find the agencies with the top mean scores
candidates2022.groupby('List Agency Desc')['Adj. FA'].mean().dropna().sort_values(asc

```

```

Out[131]: List Agency Desc
          TRIBOROUGH BRIDGE AND TUNNEL AUTHORITY 100.00
          OFFICE OF ADMINISTRATIVE TRIALS AND HEARINGS (OATH) 100.00
          LAW DEPARTMENT 100.00
          DEPARTMENT OF VETERANS' SERVICES 100.00
          NEW YORK CITY FIRE PENSION FUND 100.00
          DEPARTMENT OF PROBATION 100.00
          NYC EMPLOYEES' RETIREMENT SYSTEM 100.00
          DEPARTMENT OF INVESTIGATION 100.00
          DEPARTMENT OF HOMELESS SERVICES 100.00
          NYC POLICE PENSION FUND 100.00
          DISTRICT ATTORNEY-NEW YORK COUNTY 100.00
          DISTRICT ATTORNEY-RICHMOND COUNTY 100.00
          COMMUNITY BOARD NO.3-BRONX 100.00
          TAXI AND LIMOUSINE COMMISSION 100.00
          CITY CLERK 100.00
          TAX COMMISSION 100.00
          DEPARTMENT OF CITY PLANNING 100.00
          CIVIL SERVICE COMMISSION 100.00
          OFFICE OF MANAGEMENT AND BUDGET 100.00
          DEPARTMENT OF YOUTH AND COMMUNITY DEVELOPMENT 99.70
          N.Y.C. TRANSIT AUTHORITY 99.65
          DISTRICT ATTORNEY-BRONX COUNTY 99.60
          DEPARTMENT OF DESIGN AND CONSTRUCTION 99.43
          DEPARTMENT OF BUILDINGS 99.38
          DEPARTMENT OF ENVIRONMENTAL PROTECTION 99.28
          DEPARTMENT FOR THE AGING 99.25
          OFFICE OF THE COMPTROLLER 98.79
          DEPARTMENT OF SMALL BUSINESS SERVICES 98.75
          DEPARTMENT OF FINANCE 98.68
          DEPARTMENT OF INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS 98.65
          ADMINISTRATION FOR CHILDREN'S SERVICES 98.61
          DEPARTMENT OF TRANSPORTATION 98.59
          DEPARTMENT OF HEALTH AND MENTAL HYGIENE 98.40
          FINANCIAL INFORMATION SERVICES AGENCY 98.38
          DEPARTMENT OF CITYWIDE ADMINISTRATIVE SERVICES 98.14
          DISTRICT ATTORNEY-KINGS COUNTY 98.00
          DEPARTMENT OF EDUCATION 97.91
          DEPARTMENT OF PARKS & RECREATION 97.86
          DEPARTMENT OF CORRECTION 97.61
          DISTRICT ATTORNEY-QUEENS COUNTY 97.60
          HRA/DEPARTMENT OF SOCIAL SERVICES 97.53
          N.Y.C. HOUSING AUTHORITY 97.53
          HOUSING PRESERVATION & DEVELOPMENT 97.38
          DEPARTMENT OF CONSUMER AFFAIRS 97.00
          OFFICE OF PAYROLL ADMINISTRATION 96.58
          BOROUGH PRESIDENT-BROOKLYN 96.25
          DEPARTMENT OF SANITATION 96.16
          TEACHERS' RETIREMENT SYSTEM 96.00
          DEPT. OF RECORDS AND INFORMATION SERVICES 94.50
          CIVILIAN COMPLAINT REVIEW BOARD 91.00
          OPEN COMPETITIVE 90.79
          OFFICE OF LABOR RELATIONS 90.00
          FIRE DEPARTMENT 87.50
          POLICE DEPARTMENT 87.07
          INDEPENDENT BUDGET OFFICE 83.50
          Name: Adj. FA, dtype: float64

```

```
In [132... # find the agencies with the top median scores
```

```
candidates2022.groupby('List Agency Desc')['Adj. FA'].median().dropna().sort_values(asc
```

Out[132]:

List Agency Desc	
ADMINISTRATION FOR CHILDREN'S SERVICES	100.00
N.Y.C. TRANSIT AUTHORITY	100.00
DISTRICT ATTORNEY-BRONX COUNTY	100.00
DISTRICT ATTORNEY-KINGS COUNTY	100.00
DISTRICT ATTORNEY-NEW YORK COUNTY	100.00
DISTRICT ATTORNEY-QUEENS COUNTY	100.00
DISTRICT ATTORNEY-RICHMOND COUNTY	100.00
FINANCIAL INFORMATION SERVICES AGENCY	100.00
HOUSING PRESERVATION & DEVELOPMENT	100.00
HRA/DEPARTMENT OF SOCIAL SERVICES	100.00
LAW DEPARTMENT	100.00
N.Y.C. HOUSING AUTHORITY	100.00
NEW YORK CITY FIRE PENSION FUND	100.00
DEPARTMENT OF VETERANS' SERVICES	100.00
NYC EMPLOYEES' RETIREMENT SYSTEM	100.00
NYC POLICE PENSION FUND	100.00
OFFICE OF ADMINISTRATIVE TRIALS AND HEARINGS (OATH)	100.00
OFFICE OF LABOR RELATIONS	100.00
OFFICE OF MANAGEMENT AND BUDGET	100.00
OFFICE OF PAYROLL ADMINISTRATION	100.00
OFFICE OF THE COMPTROLLER	100.00
TAX COMMISSION	100.00
TAXI AND LIMOUSINE COMMISSION	100.00
TEACHERS' RETIREMENT SYSTEM	100.00
DEPARTMENT OF YOUTH AND COMMUNITY DEVELOPMENT	100.00
TRIBOROUGH BRIDGE AND TUNNEL AUTHORITY	100.00
DEPARTMENT OF TRANSPORTATION	100.00
DEPARTMENT OF ENVIRONMENTAL PROTECTION	100.00
CITY CLERK	100.00
CIVIL SERVICE COMMISSION	100.00
COMMUNITY BOARD NO.3-BRONX	100.00
DEPARTMENT FOR THE AGING	100.00
DEPARTMENT OF BUILDINGS	100.00
DEPARTMENT OF CITY PLANNING	100.00
DEPARTMENT OF CITYWIDE ADMINISTRATIVE SERVICES	100.00
DEPARTMENT OF CONSUMER AFFAIRS	100.00
DEPARTMENT OF CORRECTION	100.00
DEPARTMENT OF DESIGN AND CONSTRUCTION	100.00
DEPARTMENT OF EDUCATION	100.00
DEPARTMENT OF FINANCE	100.00
DEPARTMENT OF HEALTH AND MENTAL HYGIENE	100.00
DEPARTMENT OF HOMELESS SERVICES	100.00
DEPARTMENT OF INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS	100.00
DEPARTMENT OF INVESTIGATION	100.00
DEPARTMENT OF PARKS & RECREATION	100.00
DEPARTMENT OF PROBATION	100.00
DEPARTMENT OF SANITATION	100.00
DEPARTMENT OF SMALL BUSINESS SERVICES	100.00
OPEN COMPETITIVE	98.50
BOROUGH PRESIDENT-BROOKLYN	96.25
DEPT. OF RECORDS AND INFORMATION SERVICES	94.50
CIVILIAN COMPLAINT REVIEW BOARD	91.00
FIRE DEPARTMENT	88.00
INDEPENDENT BUDGET OFFICE	83.50
POLICE DEPARTMENT	82.66

Name: Adj. FA, dtype: float64

## 5. Examine combinations of variables in the dataset

### 5.1 Analysis of scores for the Fire Department only

```
In [152... candidates2022['Adj. FA'].loc[candidates2022['List Agency Desc'] == 'FIRE DEPARTMENT']
```

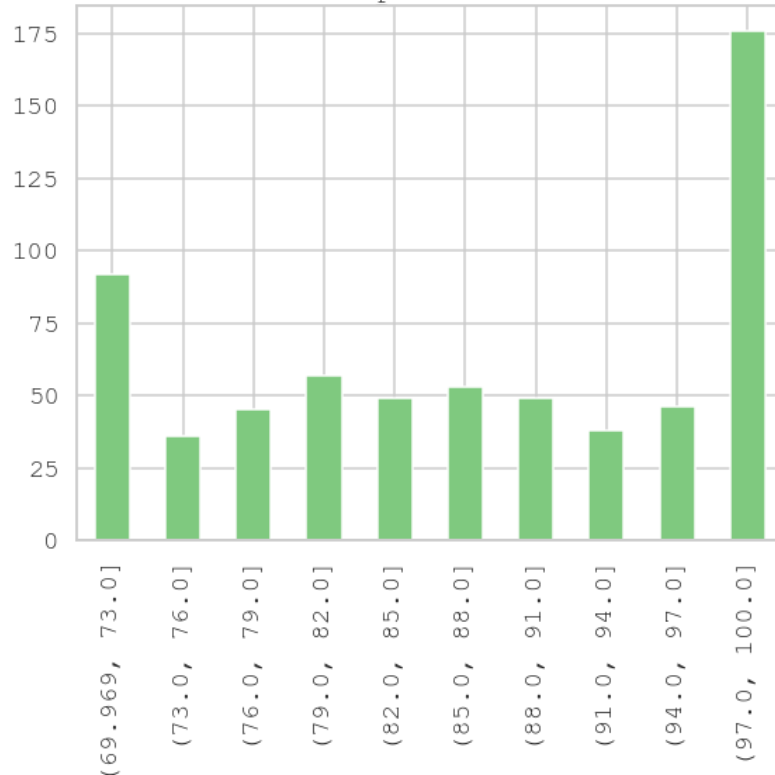
```
Out[152]: (94.0, 100.0]    222  
(69.969, 76.0]   128  
(76.0, 82.0]    102  
(82.0, 88.0]    102  
(88.0, 94.0]    87  
Name: Adj. FA, dtype: int64
```

```
In [160... candidates2022['Adj. FA'].loc[candidates2022['List Agency Desc'] == 'FIRE DEPARTMENT']
```

```
Out[160]: (94.0, 100.0]    0.35  
(69.969, 76.0]   0.20  
(76.0, 82.0]    0.16  
(82.0, 88.0]    0.16  
(88.0, 94.0]    0.14  
Name: Adj. FA, dtype: float64
```

```
In [151... candidates2022['Adj. FA'].loc[candidates2022['List Agency Desc'] == 'FIRE DEPARTMENT']
```

Distribution of 2022 NYC Fire Dept Civil Service Exam Candidate Scores



---

**Next steps**

```
In [157... # export data for data graphic creation  
firefighter_scores = candidates2022['Adj. FA'].loc[candidates2022['List Agency Desc']]
```

```
In [158... firefighter_scores.to_csv('firefighter_scores.csv')
```

---

---