

# Name Analysis of Permit Data

## Steps Done:

- 1. Import Data/modules
- 2. Filter Data by names and the alteration job type
- 3. Analyze the job-type activities of the top 10 names
- 4. Observations

## Observations

- 1. 'SINGH' was the highest name that showed up by a large margin
- 2. Although SINGH was the most occuring, they were less than the other top names for alteration A2
- 3. SINGH was leading in New Building Permits
- 4. Out of the top names, they all had very high number of A2 type jobs
  - A2 is standard interior demolition and/or renovation work that doesn't affect the use, egress or occupancy of the space.
  - <https://www.milrose.com/insights/get-to-know-the-nyc-building-application-types> (<https://www.milrose.com/insights/get-to-know-the-nyc-building-application-types>)
- 5. Name analysis helped me understand that A2 is important job type and will help me filter out for the top 10 zip codes
- 6. These could be future architect/contractors to take a look at.

## Future Work

- 1. Create criteria PER JOB TYPE and use that to filter top 10 zip codes
- 2. Find business locations of these names if they are contractors or architects
- 3. Compare zip codes to average price in the zip code to see what job prices on average are associated with these names

In [1]: `import pandas as pd`

In [4]: `data=pd.read_csv(filepath_or_buffer = '../capstone 1/DOB_Permit_Issuance.csv')`

```
In [3]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3747452 entries, 0 to 3747451
Data columns (total 60 columns):
#   Column                                Dtype
---  -
0   BOROUGH                               object
1   Bin #                                object
2   House #                              object
3   Street Name                          object
4   Job #                                int64
5   Job doc. #                           int64
6   Job Type                             object
7   Self_Cert                            object
8   Block                                object
9   Lot                                  object
10  Community Board                      object
11  Zip Code                             float64
12  Bldg Type                            float64
13  Residential                          object
14  Special District 1                   object
15  Special District 2                   object
16  Work Type                            object
17  Permit Status                        object
18  Filing Status                        object
19  Permit Type                          object
20  Permit Sequence #                    int64
21  Permit Subtype                       object
22  Oil Gas                              object
23  Site Fill                            object
24  Filing Date                          object
25  Issuance Date                        object
26  Expiration Date                      object
27  Job Start Date                       object
28  Permittee's First Name                object
29  Permittee's Last Name                object
30  Permittee's Business Name            object
31  Permittee's Phone #                  object
32  Permittee's License Type              object
33  Permittee's License #                object
34  Act as Superintendent                object
35  Permittee's Other Title               object
36  HIC License                          object
37  Site Safety Mgr's First Name          object
38  Site Safety Mgr's Last Name           object
39  Site Safety Mgr Business Name         object
40  Superintendent First & Last Name      object
41  Superintendent Business Name          object
42  Owner's Business Type                 object
43  Non-Profit                           object
44  Owner's Business Name                 object
45  Owner's First Name                    object
46  Owner's Last Name                     object
47  Owner's House #                       object
48  Owner's House Street Name             object
49  Owner's House City                    object
50  Owner's House State                   object
51  Owner's House Zip Code                 object
52  Owner's Phone #                       object
53  DOBRunDate                           object
54  PERMIT_SI_NO                          int64
55  LATITUDE                             float64
56  LONGITUDE                             float64
57  COUNCIL_DISTRICT                     float64
58  CENSUS_TRACT                         float64
59  NTA_NAME                             object
dtypes: float64(6), int64(4), object(50)
memory usage: 1.7+ GB
```

```
In [39]: data.columns = data.columns.str.replace("'", " ")
```

```
In [8]: pd.set_option('display.max_columns', None)
```

Out[8]:

	BOROUGH	Bin #	House #	Street Name	Job #	Job doc. #	Job Type	Self_Cert	Block	Lot
0	MANHATTAN	1077287	1230	6TH AVENUE	123725807	1	A2	Y	1264	5
1	STATEN ISLAND	5113169	715	OCEAN TERRACE	500876037	1	A2	Y	683	1
2	BROOKLYN	3253458	9952	3 AVE	321963014	1	DM	N	6133	56
3	BROOKLYN	3117942	179	LOTT STREET	322006618	1	DM	N	5136	58
4	BROOKLYN	3210296	2917	AVENUE N	321996970	1	DM	N	7665	4
...	...	...	...	...	...	...	...	...	...	...
3747447	QUEENS	4205475	107-24	120 STREET	421930101	1	A2	Y	9599	24
3747448	BROOKLYN	3055424	1010	BEDFORD AVENUE	310155925	3	A2	Y	1942	12
3747449	BROOKLYN	3127666	119	WEBSTER AVENUE	320267386	1	A1	N	5416	1
3747450	QUEENS	4057949	73-17	52 COURT	440626798	1	A1	N	2488	59
3747451	BROOKLYN	3008508	96	KING STREET	320911652	1	NB	N	546	42

3747452 rows × 60 columns

◀

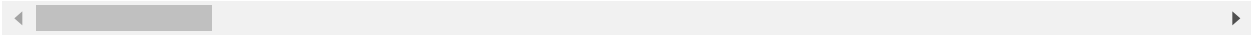
▶

In [23]:

data[(data['House #'] == "144-21") & (data['Street Name'] == "LAKEWOOD AVENUE")]

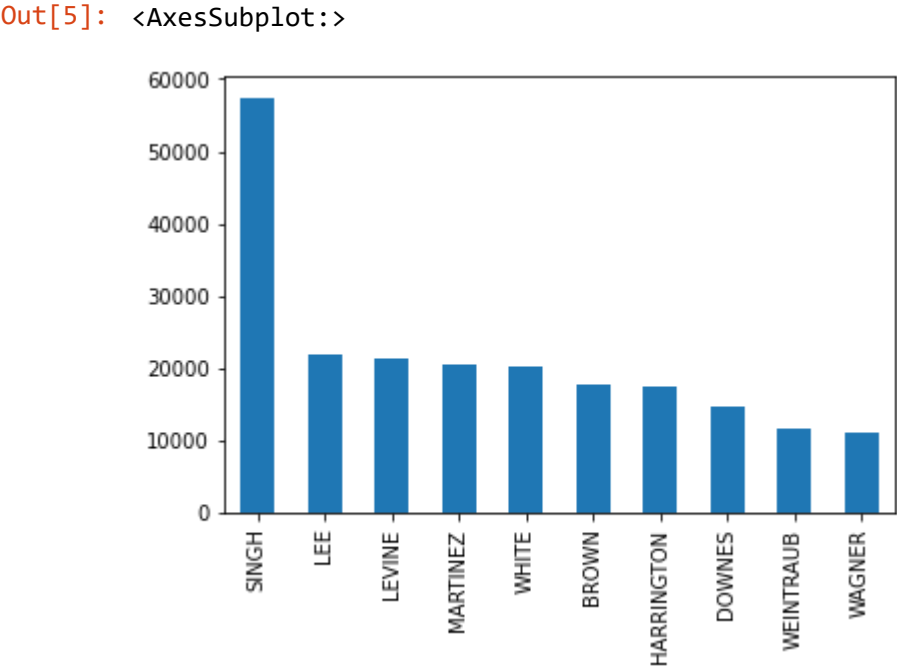
Out[23]:

	BOROUGH	Bin #	House #	Street Name	Job #	Job doc. #	Job Type	Self_Cert	Block	Lo
1116040	QUEENS	4596265	144-21	LAKEWOOD AVENUE	421056252	1	NB	N	10078	41
1411584	QUEENS	4596265	144-21	LAKEWOOD AVENUE	421056252	1	NB	N	10078	41
2519390	QUEENS	4596265	144-21	LAKEWOOD AVENUE	421056252	1	NB	N	10078	00041
2959673	QUEENS	4596265	144-21	LAKEWOOD AVENUE	421056252	1	NB	N	10078	41
3064446	QUEENS	4596265	144-21	LAKEWOOD AVENUE	421056252	1	NB	N	10078	00041
3526723	QUEENS	4596265	144-21	LAKEWOOD AVENUE	421056252	1	NB	N	10078	41
3526793	QUEENS	4596265	144-21	LAKEWOOD AVENUE	421056252	1	NB	N	10078	41
3526794	QUEENS	4596265	144-21	LAKEWOOD AVENUE	421056252	1	NB	N	10078	41
3526796	QUEENS	4596265	144-21	LAKEWOOD AVENUE	421056252	1	NB	N	10078	41
3557811	QUEENS	4596265	144-21	LAKEWOOD AVENUE	421056252	1	NB	N	10078	41



In [5]:

names = data['Permittee s Last Name'].value\_counts()  
names[:10].plot(kind='bar')



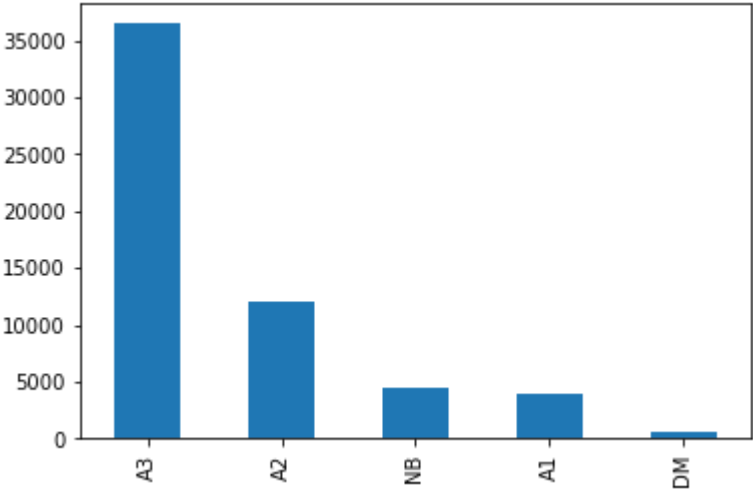
```
In [168]: data['Permittee s Last Name'].value_counts()
```

```
Out[168]: SINGH          57474
          LEE           21788
          LEVINE        21344
          MARTINEZ      20479
          WHITE         20050
          ...
          BURAY          1
          MEKULI         1
          CANRIDI        1
          PIAGIOULIATOS  1
          FRUGME         1
          Name: Permittee s Last Name, Length: 91828, dtype: int64
```

```
In [4]: data_singh = data[data['Permittee s Last Name']=='SINGH']
```

```
In [5]: data_singh_types = data_singh['Job Type'].value_counts()
data_singh_types.plot(kind='bar')
```

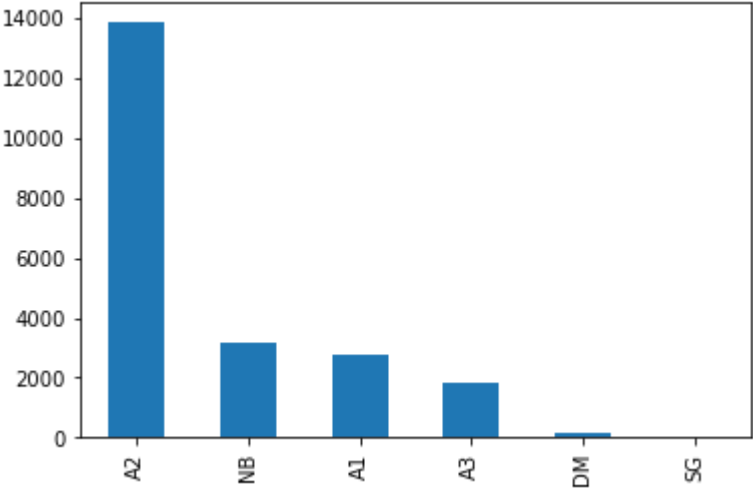
Out[5]: <AxesSubplot:>



```
In [6]: data_lee = data[data['Permittee s Last Name']=='LEE']
```

```
In [7]: data_lee_types = data_lee['Job Type'].value_counts()
data_lee_types.plot(kind='bar')
```

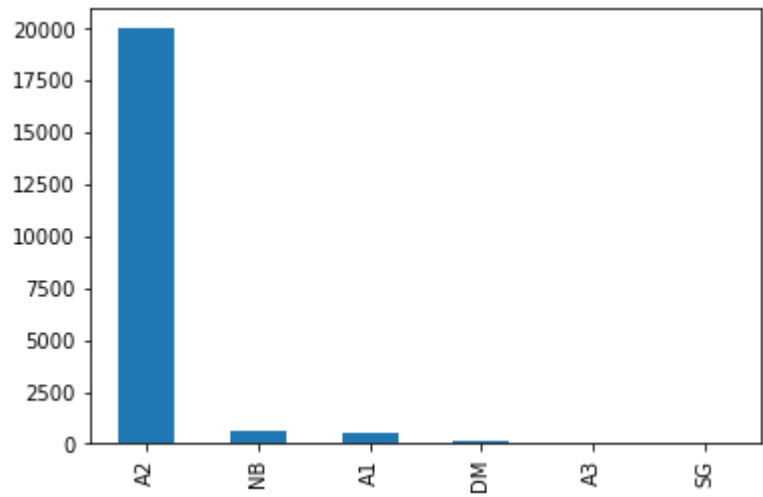
Out[7]: <AxesSubplot:>



```
In [10]: data_levine = data[data['Permittee s Last Name']=='LEVINE']
```

```
In [11]: data_levine_types = data_levine['Job Type'].value_counts()
data_levine_types.plot(kind='bar')
```

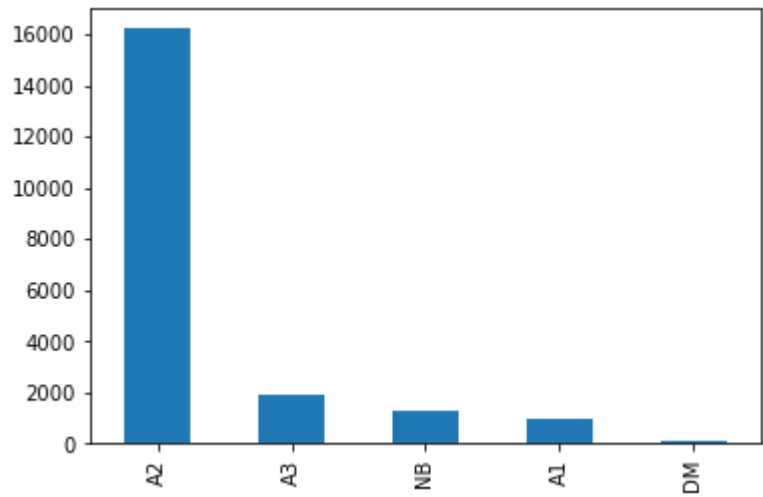
Out[11]: <AxesSubplot:>



```
In [12]: data_martinez = data[data['Permittee s Last Name']=='MARTINEZ']
```

```
In [13]: data_martinez_types = data_martinez['Job Type'].value_counts()
data_martinez_types.plot(kind='bar')
```

Out[13]: <AxesSubplot:>



```
In [40]: #Job Type = A2 and top 10 people last names
data_names_top10 = (data['Permittee s Last Name'].value_counts().index.to_list())
data_names_top10
```

Out[40]: ['SINGH',  
'LEE',  
'LEVINE',  
'MARTINEZ',  
'WHITE',  
'BROWN',  
'HARRINGTON',  
'DOWNES',  
'WEINTRAUB',  
'WAGNER']

```
In [ ]: job_types = ['A2', 'A1', 'NB', 'DM']
```

```
In [51]: df_data_names_top10=data[(data['Permittee s Last Name'].isin(data_names_top10)) &
```

```
In [52]: df_data_names_top10['Permittee s Last Name'].value_counts()
```

Out[52]:

LEVINE	21289
SINGH	21011
LEE	19913
WHITE	19640
MARTINEZ	18575
WEINTRAUB	11432
WAGNER	10868
BROWN	7277
HARRINGTON	430
DOWNES	43

Name: Permittee s Last Name, dtype: int64

```
In [54]: df_data_names_top10['Job Type'].value_counts()
```

Out[54]:

A2	95100
NB	17459
A1	16212
DM	1707

Name: Job Type, dtype: int64

```
In [136]: last_names = data_names_top10
job_types = ['A2', 'A1', 'NB', 'DM']
```

```
In [141]: df_data_names_top10_grouped = df_data_names_top10.groupby(['Permittee s Last Name', 'Job Type'])
```

```
In [142]: df11 = pd.DataFrame(df_data_names_top10_grouped.size())
```

```
In [143]: df11 = df11.set_index(['Permittee s Last Name', 'Job Type'])
# df11 = df11.set_index(['Permittee s Last Name'])
```

In [144]:

df11

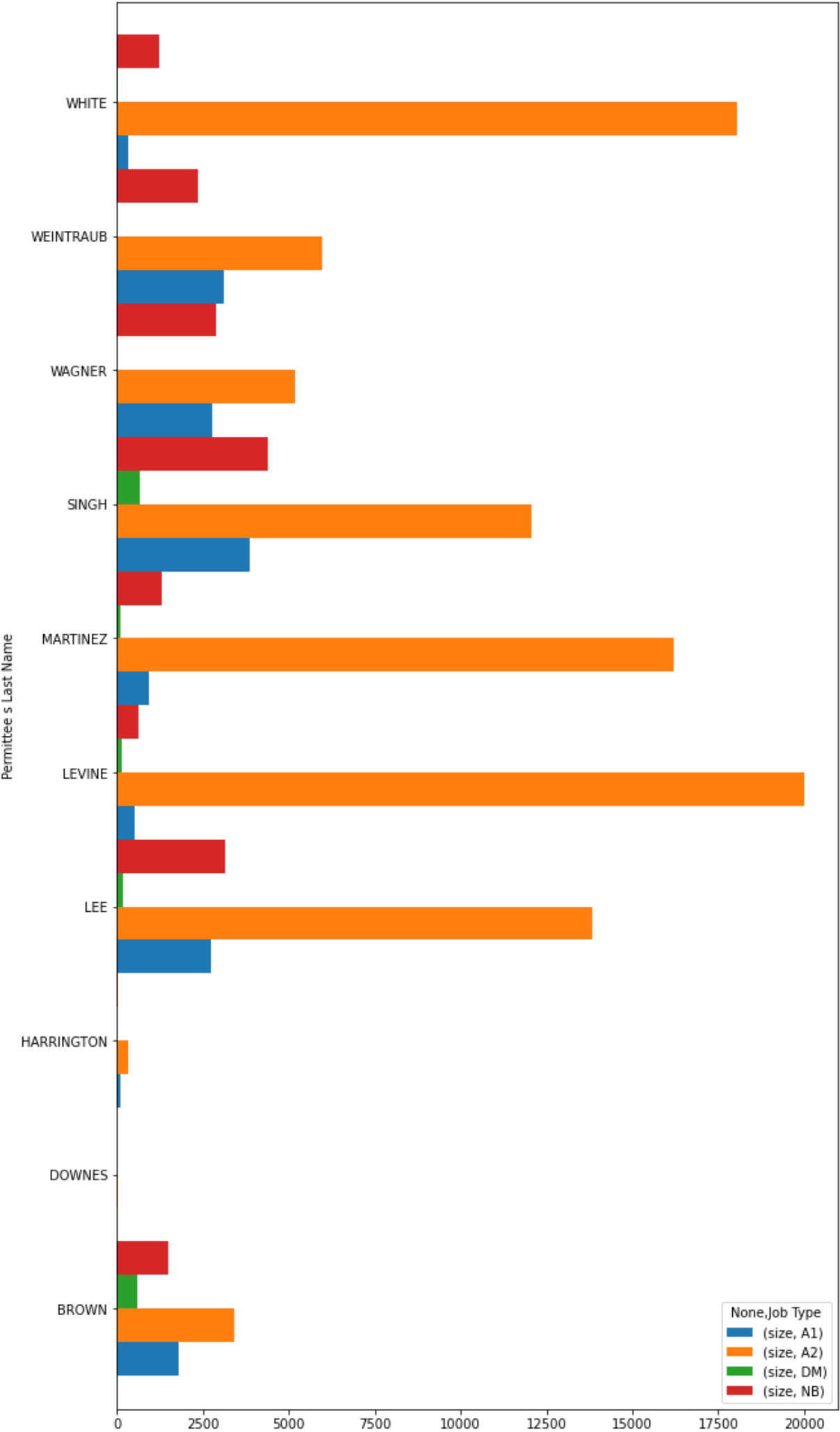
Out[144]:

Permittee s Last Name	Job Type	size
BROWN	A1	1788
	A2	3418
	DM	589
DOWNES	NB	1482
	A1	2
	A2	41
HARRINGTON	A1	102
	A2	318
	DM	1
LEE	NB	9
	A1	2750
	A2	13828
LEVINE	DM	190
	NB	3145
	A1	529
MARTINEZ	A2	19998
	DM	129
	NB	633
SINGH	A1	943
	A2	16214
	DM	109
WAGNER	NB	1309
	A1	3867
	A2	12076
WEINTRAUB	DM	670
	NB	4398
	A1	2791
WHITE	A2	5172
	DM	5
	NB	2900
	A1	3101
	A2	5974
	DM	6
	NB	2351
	A1	339
	A2	18061
	DM	8
	NB	1232



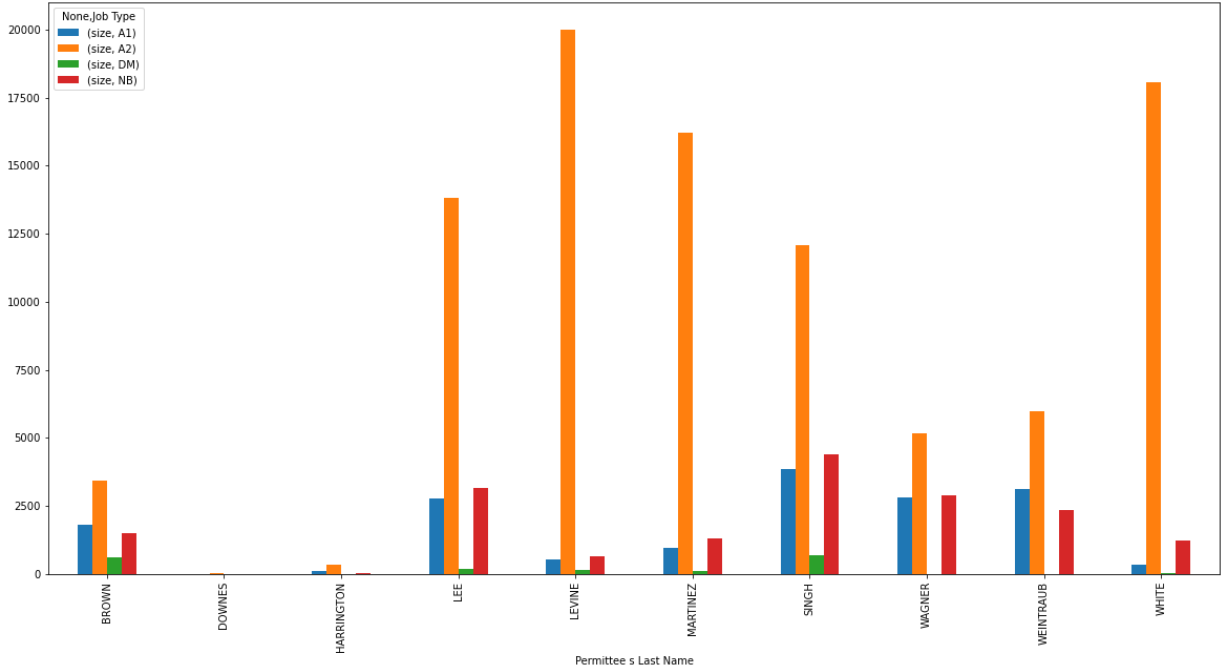
```
In [165]: df11.unstack().plot(kind='barh', figsize= (10,20), width=1)
```

Out[165]: <AxesSubplot:ylabel='Permittee s Last Name'>



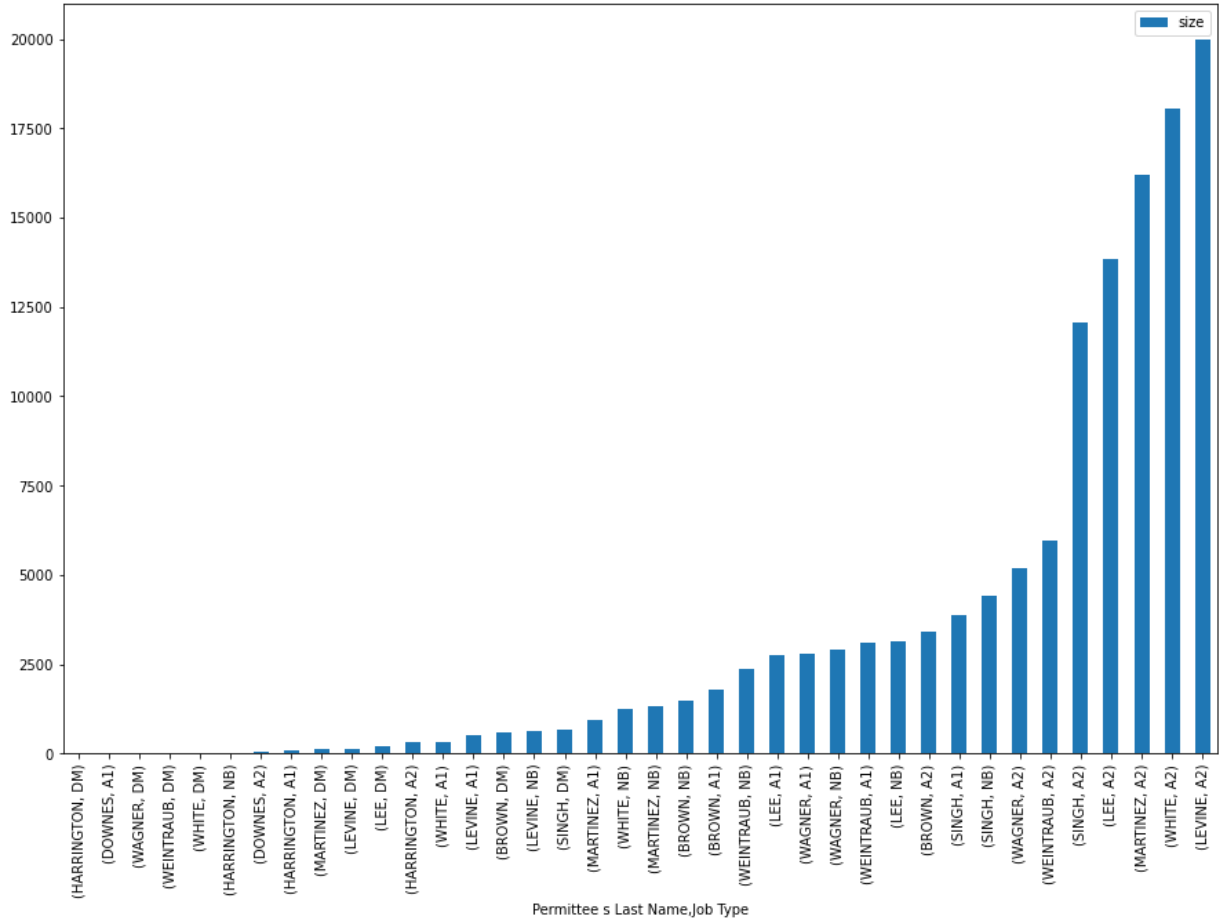
```
In [166]: df11.unstack().plot(kind='bar', figsize= (20,10))
```

Out[166]: <AxesSubplot:xlabel='Permittee s Last Name'>



```
In [159]: df11.sort_values('size').plot(kind='bar', figsize= (15,10))
```

Out[159]: <AxesSubplot:xlabel='Permittee s Last Name,Job Type'>



```
In [87]: last_names = data_names_top10
job_types = ['A2', 'A1', 'NB', 'DM']
```

Out[87]:

	Permittee s Last Name	Job Type	size
0	BROWN	A1	1788
1	BROWN	A2	3418
2	BROWN	DM	589
3	BROWN	NB	1482
4	DOWNES	A1	2
5	DOWNES	A2	41
6	HARRINGTON	A1	102
7	HARRINGTON	A2	318
8	HARRINGTON	DM	1
9	HARRINGTON	NB	9
10	LEE	A1	2750
11	LEE	A2	13828
12	LEE	DM	190
13	LEE	NB	3145
14	LEVINE	A1	529
15	LEVINE	A2	19998
16	LEVINE	DM	129
17	LEVINE	NB	633
18	MARTINEZ	A1	943
19	MARTINEZ	A2	16214
20	MARTINEZ	DM	109
21	MARTINEZ	NB	1309
22	SINGH	A1	3867
23	SINGH	A2	12076
24	SINGH	DM	670
25	SINGH	NB	4398
26	WAGNER	A1	2791
27	WAGNER	A2	5172
28	WAGNER	DM	5
29	WAGNER	NB	2900
30	WEINTRAUB	A1	3101
31	WEINTRAUB	A2	5974
32	WEINTRAUB	DM	6
33	WEINTRAUB	NB	2351
34	WHITE	A1	339
35	WHITE	A2	18061
36	WHITE	DM	8
37	WHITE	NB	1232