MVP

As MVP, I focused starting the project by simple data to detrmin the capacity of crowd in the holey musqe, And make it the base work of my project.

Steps of work:

1. Preparing the environment
    . Install Python on my machine.
    . Install Anaconda.
    . Install Jupyter Notebook.
    . Install libraries .

2. Import libraries

3. Download/Load dataset

In [31]:
```
#MVP_notebook.ipynb

# import library, pandas to read the dataset, matplotlib to visualize data, and re to clean the text
import pandas as pd
import csv

import matplotlib.pyplot as plt
import re
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
# read data stored in path 'root/Data/' using pd.read_csv, then store the data in df dataframe
df = pd.read_csv('Desktop\project\Ministry.csv')

#show first ten dataframe conte
df.head(10)
```

Out[31]:

	RESERVATION_NO	Permit_No	DATE_	TIME_FROM	TIME_TO	SEX_	AGE_	assembly_POINT	GATE_INTR	Permit_KIND
0	47851116	942816711	9/6/2021	6:00 PM	9:00 PM	MALE	55	موقف الجمرات	باب الملك فهد	أداء عمرة
1	47851116	942816712	9/6/2021	6:00 PM	9:00 PM	FEMALE	22	موقف الجمرات	باب الملك فهد	أداء عمرة
2	47851116	942816713	9/6/2021	6:00 PM	9:00 PM	FEMALE	25	موقف الجمرات	باب الملك فهد	أداء عمرة
3	47851117	942816714	9/6/2021	6:00 PM	9:00 PM	FEMALE	25	موقف الجمرات	باب الملك فهد	أداء عمرة
4	47851118	942816715	9/6/2021	6:00 PM	9:00 PM	FEMALE	32	الشبيكه	باب الملك فهد	أداء عمرة
5	47851119	942816716	9/6/2021	6:00 PM	9:00 PM	MALE	55	الشبيكه	باب الملك فهد	أداء عمرة
6	47851120	942816717	9/6/2021	6:00 PM	9:00 PM	MALE	22	الشبيكه	باب الملك فهد	أداء عمرة
7	47851121	942816718	9/6/2021	6:00 PM	9:00 PM	MALE	40	الشبيكه	باب الملك فهد	أداء عمرة
8	47851122	942816719	9/6/2021	6:00 PM	9:00 PM	MALE	46	الشبيكه	باب الملك فهد	أداء عمرة
9	47851123	942816720	9/6/2021	6:00 PM	9:00 PM	MALE	44	الشبيكه	باب الملك فهد	أداء عمرة

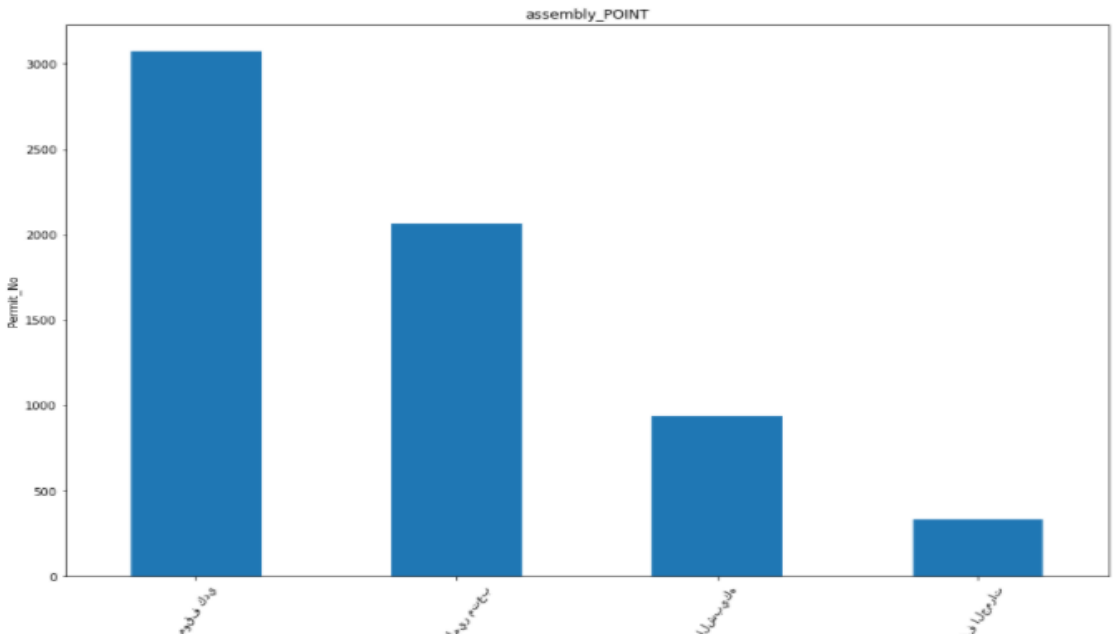
Figure (1)

4- explore the data type at each series & check to null values

```
In [32]: # explore the data type at each series
         df.info()

         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 6415 entries, 0 to 6414
         Data columns (total 10 columns):
          #   Column          Non-Null Count  Dtype
         ---  ------          --------------  -----
          0   RESERVATION_NO  6415 non-null   int64
          1   Permit_No       6415 non-null   int64
          2   DATE_           6415 non-null   object
          3   TIME_FROM       6415 non-null   object
          4   TIME_TO         6415 non-null   object
          5   SEX_            6415 non-null   object
          6   AGE_            6415 non-null   int64
          7   assembly_POINT  6415 non-null   object
          8   GATE_INTR       6415 non-null   object
          9   Permit_KIND     6415 non-null   object
         dtypes: int64(3), object(7)
         memory usage: 501.3+ KB

In [33]: # check to null values
         df.isna().sum()

Out[33]: RESERVATION_NO    0
         Permit_No         0
         DATE_             0
         TIME_FROM         0
         TIME_TO           0
         SEX_              0
         AGE_              0
         assembly_POINT    0
         GATE_INTR         0
         Permit_KIND       0
         dtype: int64

In [34]:
```

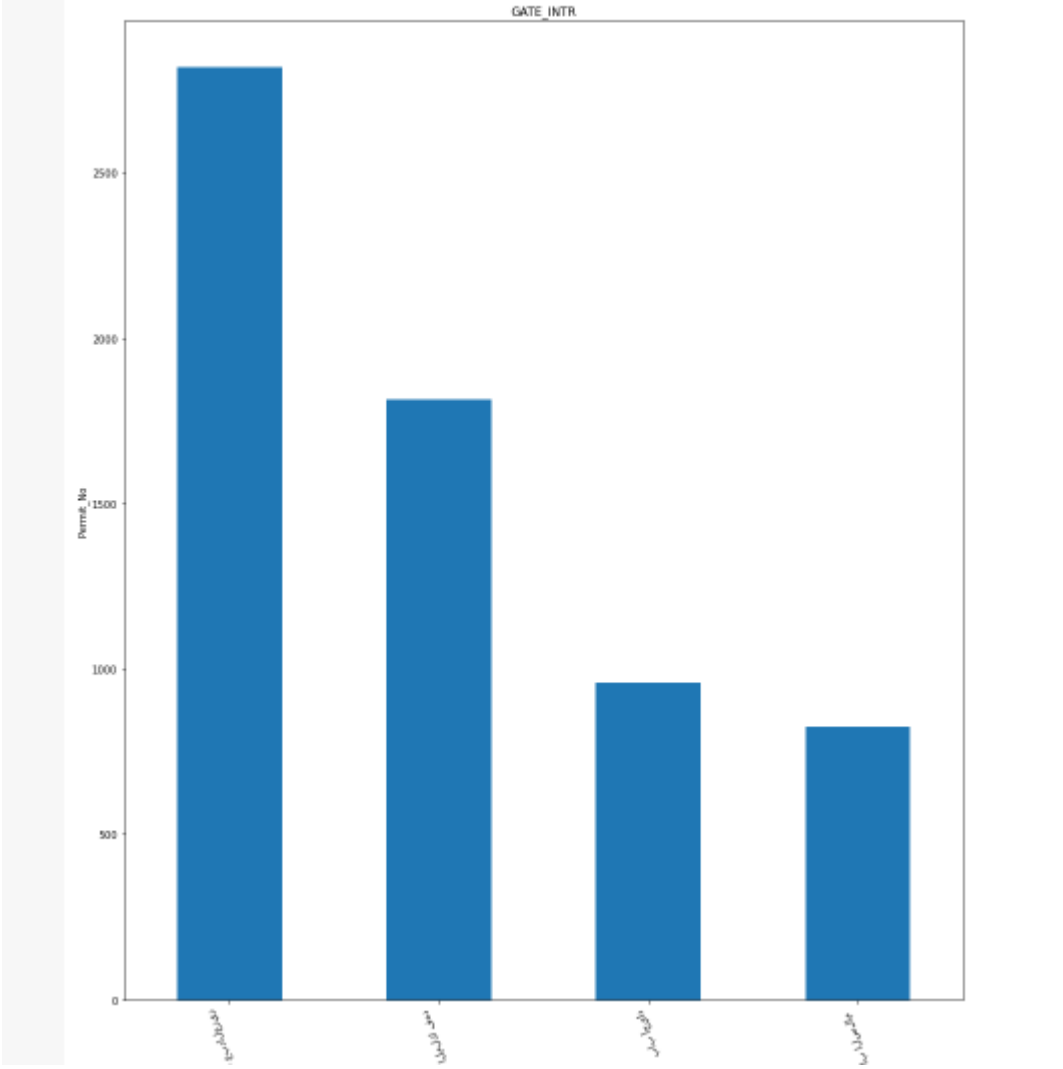
6- check which assembol_point we well focus on it.

```
In [44]: plt.figure(figsize=(15,10))
         df['assembly_POINT'].value_counts().nlargest(20).plot(kind='bar')
         plt.xticks(rotation=60)
         plt.title('assembly_POINT')
         plt.xlabel('assembly_POINT')
         plt.ylabel('Permit_No');
```



7-6- check which gate_inter we well focus on it.

```
In [47]: plt.figure(figsize=(15,18))
         df['GATE_INTR'].value_counts().nlargest(30).plot(kind='bar')
         plt.xticks(rotation=110)
         plt.title('GATE_INTR')
         plt.xlabel('GATE_INTR')
         plt.ylabel('Permit_No');
```



And that`s it for now! And we well determine the capacity for each gate_inter>