Assignement

Vivek Kumar Rai
11915876
CAP 776
B-65

In [4]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Read CSV File

In [7]:

```
df = pd.read_csv("Covid.csv")
```

In [6]:

```
df.head()
```

Out[6]:

	Unnamed: 0	1	Patient Number	State Patient Number	Date Announced	Age Bracket	Gender	Detected City	Detected District	De
0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	1	2.0	1.0	KL-TS-P1	30/01/2020	20	F	Thrissur	Thrissur	
2	2	3.0	2.0	KL-AL-P1	02/02/2020	NaN	NaN	Alappuzha	Alappuzha	
3	3	4.0	3.0	KL-KS-P1	03/02/2020	NaN	NaN	Kasaragod	Kasaragod	
4	4	5.0	4.0	DL-P1	02/03/2020	45	M	East Delhi (Mayur Vihar)	East Delhi	

5 rows × 27 columns

Extract Gender , State and age column from dataset

In [8]:

```
sub_df = df[["Gender","Detected State","Age Bracket"]]
```

In [9]:

```
sub_df.head()
```

Out[9]:

	Gender	Detected State	Age Bracket
0	NaN	NaN	NaN
1	F	Kerala	20
2	NaN	Kerala	NaN
3	NaN	Kerala	NaN
4	M	Delhi	45

In [10]:

```
sub_df.shape
```

Out[10]:

```
(2072, 3)
```

Drop NA Values

In [11]:

```
sub_df.dropna(subset = ['Detected State','Gender'],inplace=True)
```

```
C:\Users\Vineet Rai\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: Set
tingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (http://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
rsus-a-copy)
  """Entry point for launching an IPython kernel.
```

In [12]:

```
#sub_df["Gender"] = df["Gender"].apply(Lamba x:1 if x == "M" else 0)
```

In [13]:

```
sub_df.head()
```

Out[13]:

	Gender	Detected State	Age Bracket
1	F	Kerala	20
4	M	Delhi	45
5	M	Telangana	24
6	M	Rajasthan	69
21	F	Rajasthan	70

In [14]:

```
sub_df["Age Bracket"].fillna('0',inplace =True)
```

```
C:\Users\Vineet Rai\Anaconda3\lib\site-packages\pandas\core\generic.py:6287:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (http://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
rsus-a-copy)
  self._update_inplace(new_data)
```

In [15]:

```
def assign_age(x):
    if len(x)>3:
        x1,x2 = x.split('-')
        return str((int(x1)+int(x2))/2).split('.')[0]
    else:
        return x
```

In [16]:

```
sub_df['Age Bracket'] = sub_df["Age Bracket"].apply(assign_age)
```

```
C:\Users\Vineet Rai\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: Set
tingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (http://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
rsus-a-copy)
  """Entry point for launching an IPython kernel.
```

Change Age column to int type and replace NA values in Age coloumn with mean

In [17]:

```
sub_df["Age Bracket"] = sub_df['Age Bracket'].astype(int)
```

```
C:\Users\Vineet Rai\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: Set
tingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (http://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
rsus-a-copy)
  """Entry point for launching an IPython kernel.
```

In [18]:

```
sub_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 742 entries, 1 to 1663
Data columns (total 3 columns):
Gender            742 non-null object
Detected State    742 non-null object
Age Bracket       742 non-null int32
dtypes: int32(1), object(2)
memory usage: 20.3+ KB
```

In [19]:

```
age_mean = int(sub_df[sub_df["Age Bracket"]!=0]['Age Bracket'].mean())
```

In [20]:

```
sub_df["Age Bracket"].replace(0,age_mean,inplace=True)
```

```
C:\Users\Vineet Rai\Anaconda3\lib\site-packages\pandas\core\generic.py:6786:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (http://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
rsus-a-copy)
  self._update_inplace(new_data)
```

In [21]:

```
sub_df.head()
```

Out[21]:

	Gender	Detected State	Age Bracket
1	F	Kerala	20
4	M	Delhi	45
5	M	Telangana	24
6	M	Rajasthan	69
21	F	Rajasthan	70

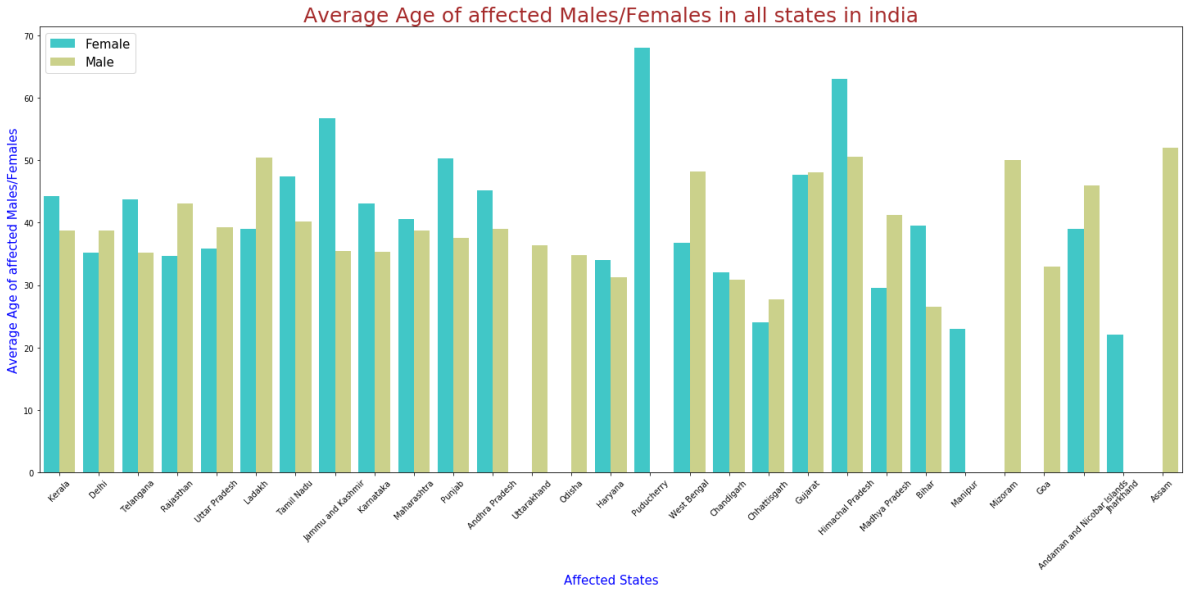
Visualization of data

In [40]:

```
plt.figure(figsize=(25,10))
plt.xticks(rotation=45)
sns.barplot(x="Detected State",y="Age Bracket",data=sub_df,hue='Gender',ci=None,
palette="rainbow")
plt.legend(["Female","Male"],prop={'size':15})
plt.xlabel("Affected States",fontdict={"Size":15},color="blue")
plt.ylabel("Average Age of affected Males/Females",fontdict={"size":15},color="blue")
plt.title("Average Age of affected Males/Females in all states in india",
fontdict={"size":25},color="brown")
#Text(0.5, 1.0, 'Average Age of affected Males/Females in all states in india')
```

Out[40]:

```
Text(0.5, 1.0, 'Average Age of affected Males/Females in all states in indi
a')
```



This barplot shows the largest age affected female in punduchery (68) and male in assam (50).
Average age of affected males and females near about 45.

In kerala we can see mostly affected females are around 45 years amd males are around 38 year so
we can say affected males are younger compartively females.

In [ ]: