```
In [1]: #import libraries
        import pandas as pd
        import plotly.express as px
        import matplotlib.pyplot as plt
```

```
In [2]: #load dataset
```

```
In [3]: dataset=pd.read_csv("preprocessed_data.csv")
        dataset
```

Out[3]:

	Education	JoiningYear	City	PaymentTier	Age	Gender	EverBenched	ExperienceInC
0	Bachelors	2017	Bangalore	3	34	Male	No	
1	Bachelors	2013	Pune	1	28	Female	No	
2	Bachelors	2014	New Delhi	3	38	Female	No	
3	Masters	2016	Bangalore	3	27	Male	No	
4	Masters	2017	Pune	3	24	Male	Yes	
...	...	...	...	...	...	...	...	
4648	Bachelors	2013	Bangalore	3	26	Female	No	
4649	Masters	2013	Pune	2	37	Male	No	
4650	Masters	2018	New Delhi	3	27	Male	No	
4651	Bachelors	2012	Bangalore	3	30	Male	Yes	
4652	Bachelors	2015	Bangalore	3	33	Male	Yes	

4653 rows × 9 columns

```
In [4]: #First we check null values
        dataset.isnull().sum()
```

```
Out[4]: Education                   0
        JoiningYear                 0
        City                        0
        PaymentTier                 0
        Age                         0
        Gender                      0
        EverBenched                 0
        ExperienceInCurrentDomain   0
        LeaveOrNot                  0
        dtype: int64
```

```
In [5]: dataset.shape
```

```
Out[5]: (4653, 9)
```

In [6]:
```python
dataset['Gender'].value_counts()['Male']
```

Out[6]: 2778

In [7]:
```python
dataset['Gender'].value_counts()['Female']
```

Out[7]: 1875

In [8]:
```python
dataset['City'].value_counts()
```

Out[8]:
```
Bangalore    2228
Pune         1268
New Delhi    1157
Name: City, dtype: int64
```

In [9]:
```python
df= dataset[dataset['Gender'] == 'Male']
df
```

Out[9]:

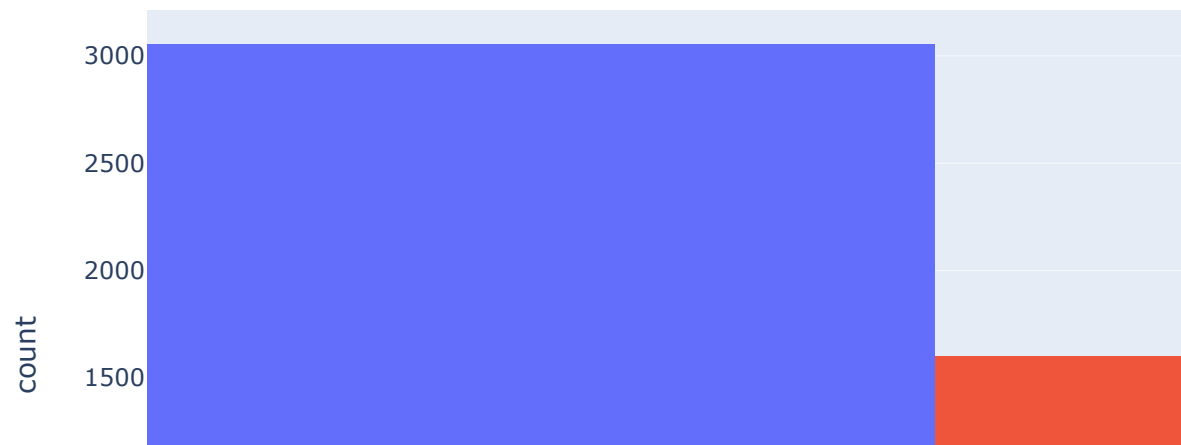|	Education	JoiningYear	City	PaymentTier	Age	Gender	EverBenched	ExperienceInC
0	Bachelors	2017	Bangalore	3	34	Male	No	
3	Masters	2016	Bangalore	3	27	Male	No	
4	Masters	2017	Pune	3	24	Male	Yes	
5	Bachelors	2016	Bangalore	3	22	Male	No	
6	Bachelors	2015	New Delhi	3	38	Male	No	
...	...	...	...	...	...	...	...	
4647	Bachelors	2016	Pune	3	30	Male	No	
4649	Masters	2013	Pune	2	37	Male	No	
4650	Masters	2018	New Delhi	3	27	Male	No	
4651	Bachelors	2012	Bangalore	3	30	Male	Yes	
4652	Bachelors	2015	Bangalore	3	33	Male	Yes	

2778 rows × 9 columns

In [10]:
```python
dataset.columns
```

Out[10]:
```
Index(['Education', 'JoiningYear', 'City', 'PaymentTier', 'Age', 'Gender',
       'EverBenched', 'ExperienceInCurrentDomain', 'LeaveOrNot'],
      dtype='object')
```

Working and Non Working Cantidate details

In [11]:
```python
fig = px.histogram(dataset, x='LeaveOrNot', color='LeaveOrNot',
 title= 'Working and Non Working Cantidate',
 labels={'TARGET':'LeaveOrNot'})
fig.show()
```
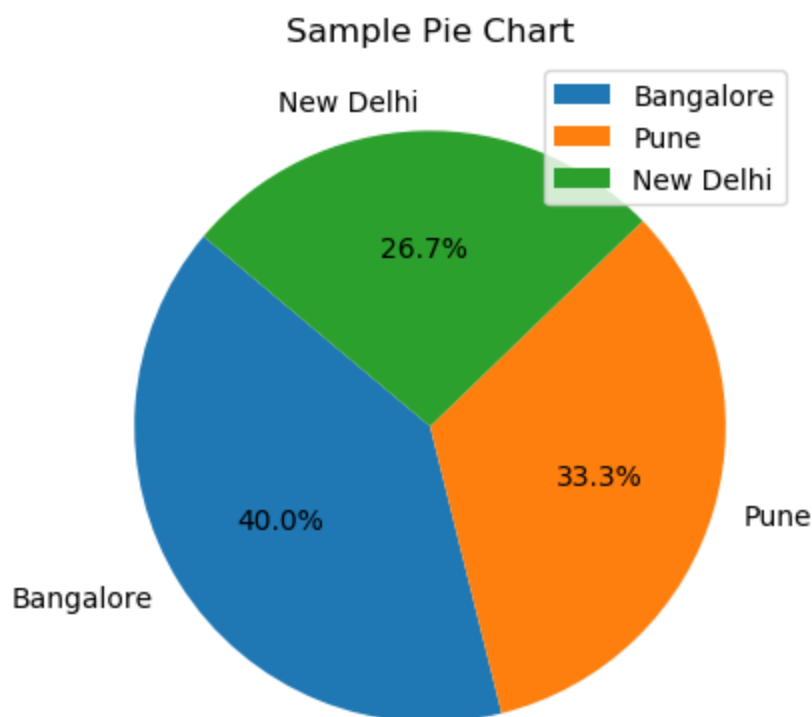
Working and Non Working Cantidate

Percentage of Employes cityWise

In [12]:
```python
labels = ['Bangalore', 'Pune', 'New Delhi']
sizes = [30, 25, 20]

plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=140)
plt.title('Sample Pie Chart')
plt.legend(labels, loc='upper right')
plt.show()
```


What kind of relation between JoiningYear and Age
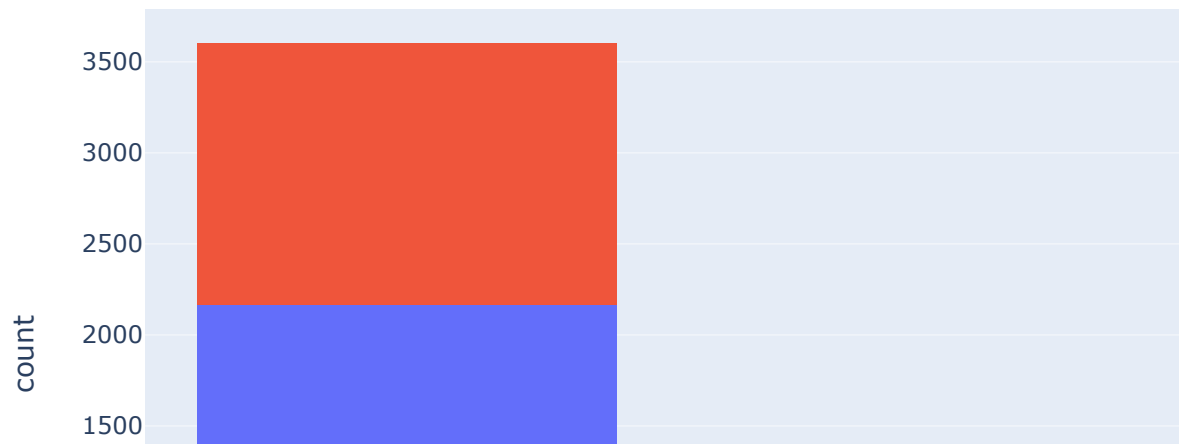
In [13]:
```python
dataset.corr()
```

Out[13]:

	JoiningYear	PaymentTier	Age	ExperienceInCurrentDomain	Lea
JoiningYear	1.000000	-0.096078	0.013165	-0.036525	(
PaymentTier	-0.096078	1.000000	0.007631	0.018314	-(
Age	0.013165	0.007631	1.000000	-0.134643	-
ExperienceInCurrentDomain	-0.036525	0.018314	-0.134643	1.000000	-(
LeaveOrNot	0.181705	-0.197638	-0.051126	-0.030504	

Relationship between joining year and age is possitive correlation but its a poor possitive value 0.013165

In [14]:
```python
fig = px.histogram(dataset, x='Education', color='Gender',
 title= 'No of Employes in Education wise').update_xaxes(categoryorder="total (
fig.show()
```
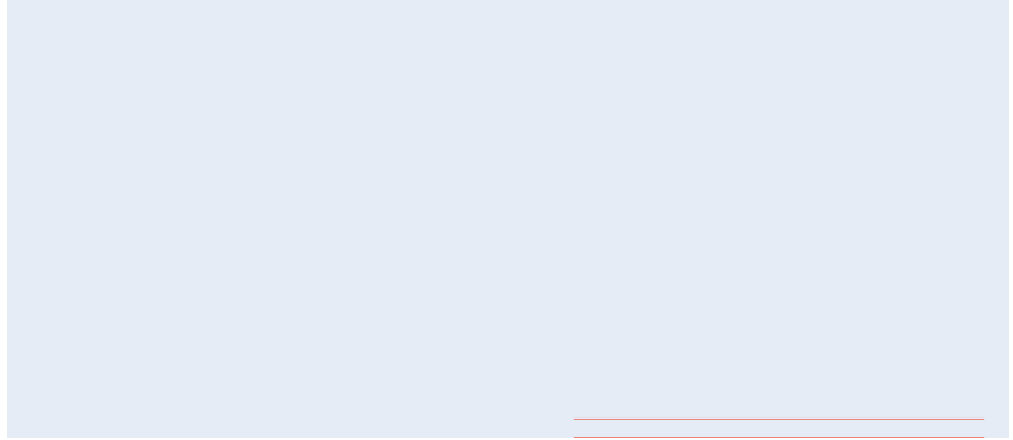
No of Employes in Education wise

In [15]:
```python
fig = px.bar(dataset, x='City', y='Education', color='Gender', title= 'citywise
fig.show()
```

citywise education

Education

In [ ]: