

```
In [3]: import pandas as pd  
df=pd.read_csv("C:/Users/Atwongire Vianney/Desktop/douglasoft/jobs_in_data.csv")
```

```
In [4]: df.head()
```

Out[4]:

	work_year	job_title	job_category	salary_currency	salary	salary_in_usd	experience_level
0	2023	Data DevOps Engineer	Data Engineering	EUR	88000	95012	Mid-level
1	2023	Data Architect	Data Architecture and Modeling	USD	186000	186000	Senior
2	2023	Data Architect	Data Architecture and Modeling	USD	81800	81800	Senior
3	2023	Data Scientist	Data Science and Research	USD	212000	212000	Senior
4	2023	Data Scientist	Data Science and Research	USD	93300	93300	Senior

```
In [15]: senior_df=df[df['experience_level']=='Senior']
```

In [16]: senior_df

Out[16]:

	work_year	job_title	job_category	salary_currency	salary	salary_in_usd	experience_level
1	2023	Data Architect	Data Architecture and Modeling	USD	186000	186000	Senio
2	2023	Data Architect	Data Architecture and Modeling	USD	81800	81800	Senio
3	2023	Data Scientist	Data Science and Research	USD	212000	212000	Senio
4	2023	Data Scientist	Data Science and Research	USD	93300	93300	Senio
5	2023	Data Scientist	Data Science and Research	USD	130000	130000	Senio
...
9341	2020	Data Science Manager	Data Science and Research	USD	190200	190200	Senio
9345	2020	Machine Learning Engineer	Machine Learning and AI	EUR	40000	45618	Senio
9346	2021	Director of Data Science	Data Science and Research	USD	168000	168000	Senio
9350	2021	Data Specialist	Data Management and Strategy	USD	165000	165000	Senio
9351	2020	Data Scientist	Data Science and Research	USD	412000	412000	Senio

6709 rows × 7 columns



In [18]: salary_in_usd_list=senior_df['salary_in_usd'].tolist()
job_category_list=senior_df['job_category'].tolist()

```
In [19]: salary_in_usd_list
```

```
Out[19]: [186000,
          81800,
          212000,
          93300,
          130000,
          100000,
          224400,
          138700,
          300000,
          234000,
          266500,
          152000,
          273400,
          182200,
          167500,
          106500,
          185900,
          129300,
          122000,
          241500]
```

```
In [20]: job_category_list
```

```
Machine Learning and AI ,
'Data Engineering',
'Data Engineering',
'Data Management and Strategy',
'Data Management and Strategy',
'Data Engineering',
'Data Engineering',
'BI and Visualization',
'BI and Visualization',
'Data Analysis',
'Data Analysis',
'Data Analysis',
'Data Analysis',
'Data Engineering',
'Data Engineering',
'BI and Visualization',
'BI and Visualization',
'Data Science and Research',
'Data Science and Research',
'Data Science and Research',
```

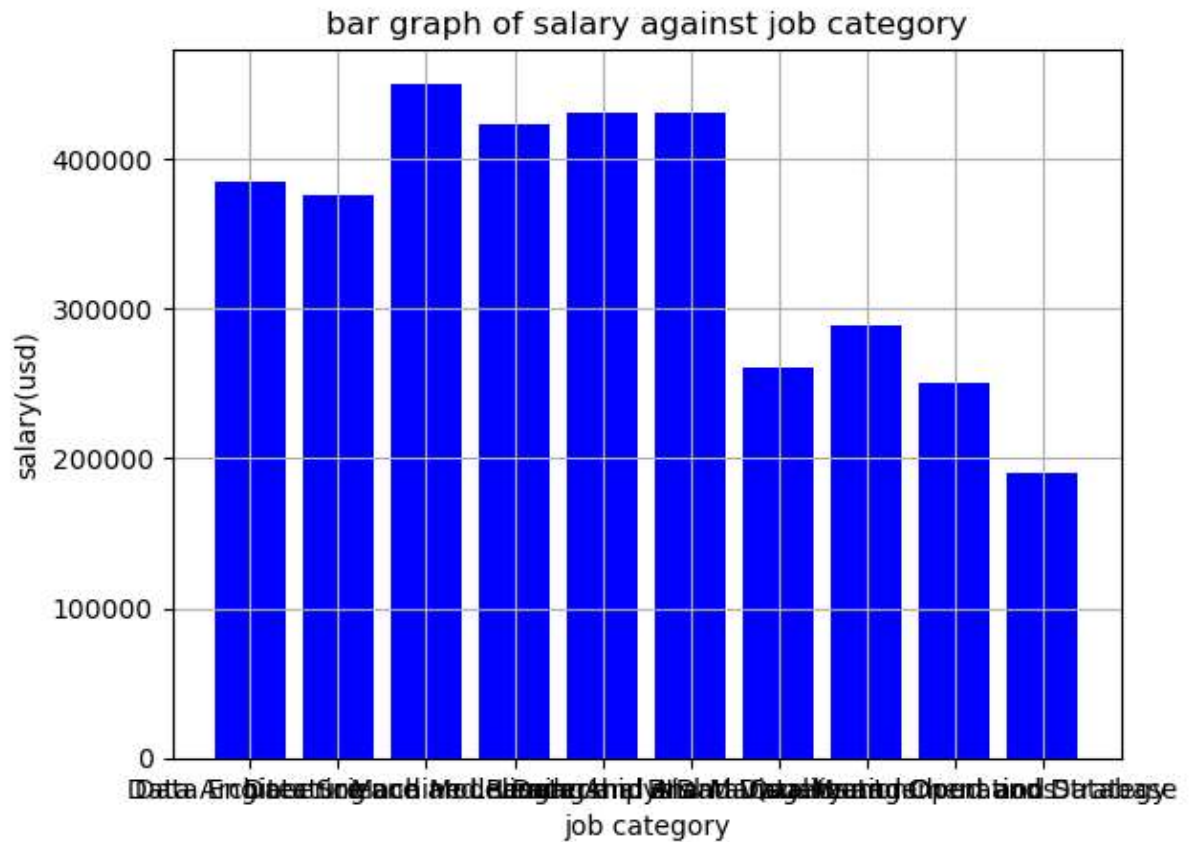
```
In [27]: import numpy as np
         #xs=np.array(_list, dtype= np.float64)
         #ys=np.array(_list, dtype= np.string64)
```

```
In [28]: xs
```

```
Out[28]: array([186000.,  81800., 212000., ..., 168000., 165000., 412000.])
```

```
In [37]: import matplotlib.pyplot as plt
plt.bar(df['job_category'],df['salary_in_usd'], color='blue')
plt.xlabel('job category')
plt.ylabel('salary(usd)')
plt.title('bar graph of salary against job category')
plt.grid(True)
plt.show
```

```
Out[37]: <function matplotlib.pyplot.show(close=None, block=None)>
```



```
In [ ]:
```