Name: Shivam Bhatt

Enrollment No: 92301733046

```
!pip install kaggle
import os
os.environ['KAGGLE_CONFIG_DIR']="/content"
!kaggle datasets download stackoverflow/stack-overflow-2018-developer-survey
!unzip /content/stack-overflow-2018-developer-survey.zip
#import basic liabraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
data2=pd.read_csv("/content/survey_results_public.csv")
data3=pd.read_csv("/content/survey_results_schema.csv")
data2.shape
```

output:

```
data2.shape
(98855, 129)
```

data3.shape

output:

```
(129, 2)
```

data3.head()

	Column	QuestionText
0	Respondent	Randomized respondent ID number (not in order ...
1	Hobby	Do you code as a hobby?
2	OpenSource	Do you contribute to open source projects?
3	Country	In which country do you currently reside?
4	Student	Are you currently enrolled in a formal, degree...

#count the number of null values in each feature

data=pd.read_csv("/content/survey_results_public.csv")

print(data)

data.isnull().sum()

```
#count the number of null values in each feature
data.isnull().sum()
```

	0
Respondent	0
Hobby	0
OpenSource	0
Country	0
Student	539
...	...
Age	13449
Dependents	14348
MilitaryUS	37205
SurveyTooLong	12922
SurveyEasy	12962

129 rows × 1 columns

dtype: int64

#count the percentage of null values

data.isnull().sum()/data.shape[0]*100

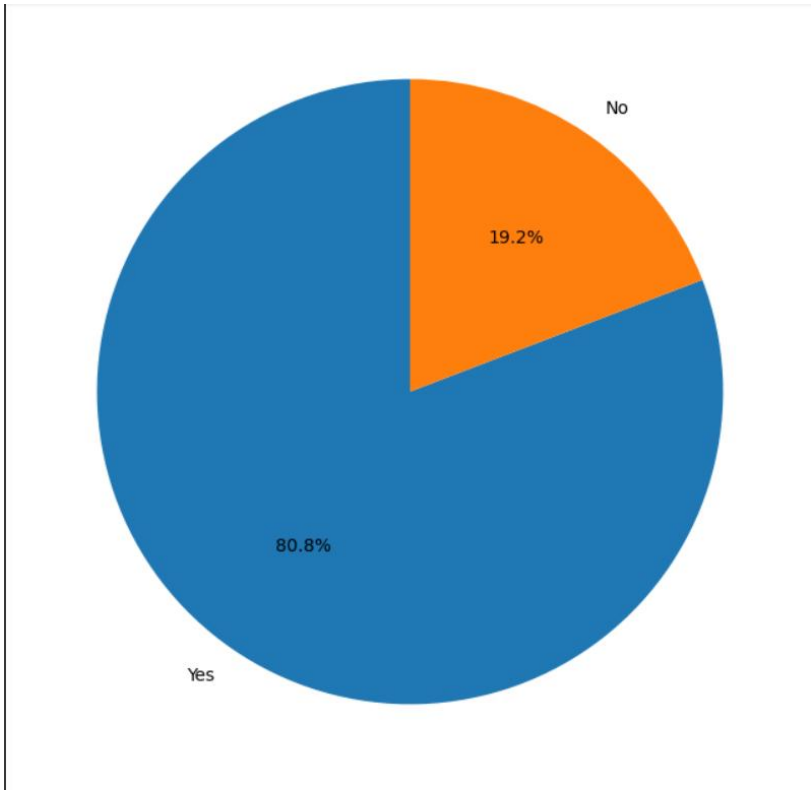
	0
Respondent	0.000000
Hobby	0.000000
OpenSource	0.000000
Country	0.000000
Student	1.184850
...	...
Age	29.564090
Dependents	31.540305
MilitaryUS	81.785408
SurveyTooLong	28.405619
SurveyEasy	28.493548

129 rows × 1 columns

dtype: float64

#draw the pychart for number of people who finds coding as hobby

```
hobby_counts = data2['Hobby'].value_counts()

plt.figure(figsize=(8, 8))
plt.pie(hobby_counts, labels=hobby_counts.index, autopct='%1.1f%%', startangle=90)
plt.title('Distribution of people who code as a hobby')
plt.show()
```



#determine the number of people contributing to open source projects

open_source_contributors = data2[data2['OpenSource'] == 'Yes'].shape[0]

print(open_source_contributors)

```
#determine the number of people contributing to open source projects
open_source_contributors = data2[data2['OpenSource'] == 'Yes'].shape[0]
print(open_source_contributors)
```

```
43086
```

#determine the top 20 countries where the responses are obtained

country_counts = data2['Country'].value_counts()

top_20_countries = country_counts.head(20)

print("Top 20 countries with the most responses:")

print(top_20_countries)

```
country_counts = data2['Country'].value_counts()
top_20_countries = country_counts.head(20)
print("Top 20 countries with the most responses:")
print(top_20_countries)
```

```
Top 20 countries with the most responses:
Country
United States          20309
India                  13721
Germany                 6459
United Kingdom          6221
Canada                  3393
Russian Federation      2869
France                  2572
Brazil                  2505
Poland                  2122
Australia               2018
Netherlands             1841
Spain                   1769
Italy                   1535
Ukraine                 1279
Sweden                  1164
Pakistan                1050
China                   1037
Switzerland             1010
Turkey                  1004
Israel                  1003
Name: count, dtype: int64
```

#do other 5 analysis as per your own thinking (which involves different charts and graphs)

1. Distribution of 'Student' status

plt.figure(figsize=(8, 6))

sns.countplot(data=data2, x='Student')

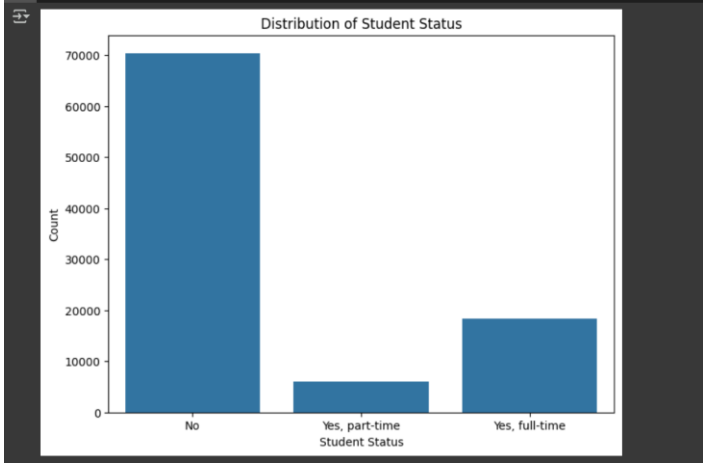
plt.title('Distribution of Student Status')

plt.xlabel('Student Status')

plt.ylabel('Count')

plt.show()

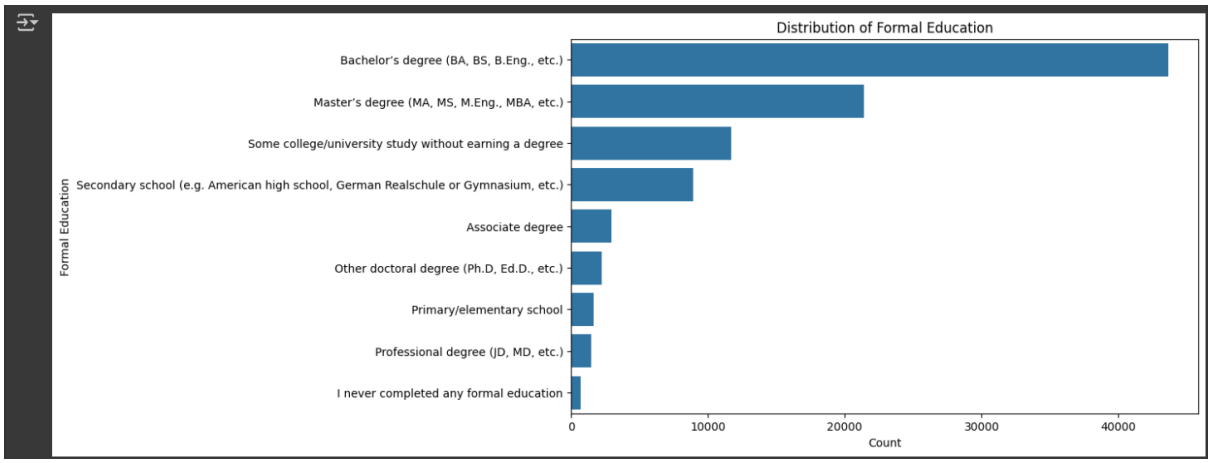
```
plt.show()
```



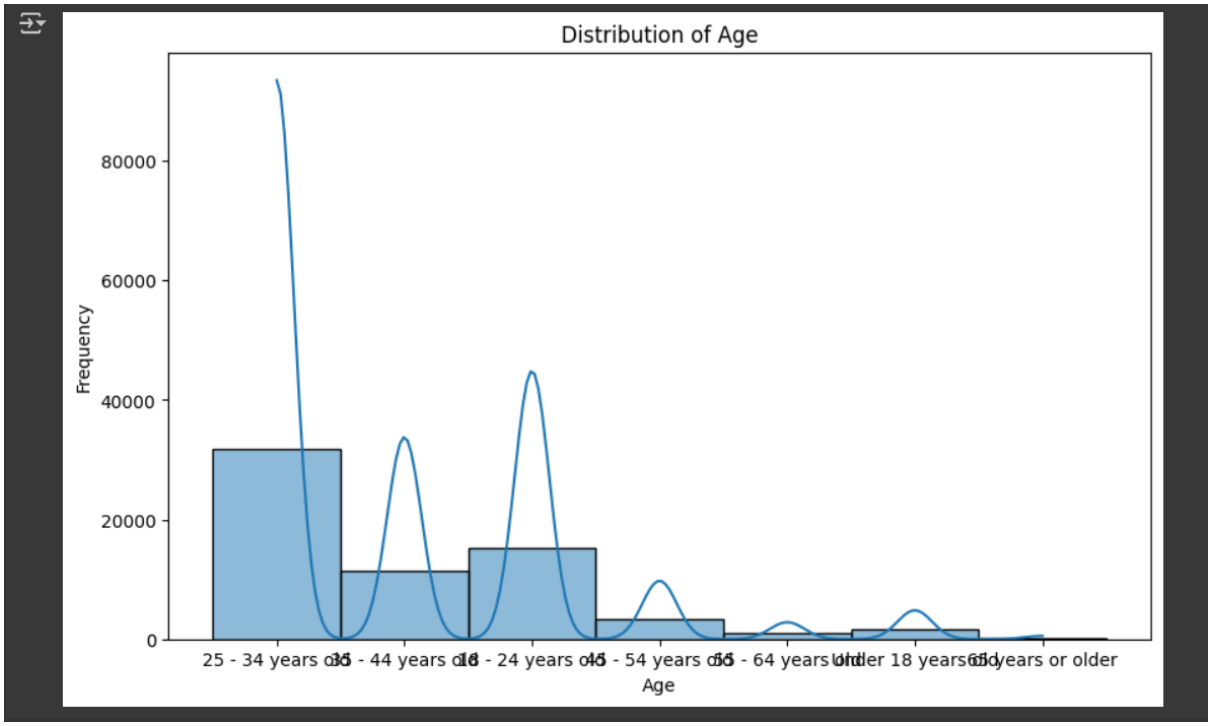
2. Distribution of 'FormalEducation'

plt.figure(figsize=(10, 6))

```
sns.countplot(data=data2, y='FormalEducation',
order=data2['FormalEducation'].value_counts().index)

plt.title('Distribution of Formal Education')

plt.xlabel('Count')

plt.ylabel('Formal Education')

plt.show()
```




```
# 3. Distribution of 'Age'

plt.figure(figsize=(10, 6))

sns.histplot(data=data2, x='Age', kde=True)

plt.title('Distribution of Age')

plt.xlabel('Age')

plt.ylabel('Frequency')

plt.show()
```



4. Distribution of 'DevType' (Developer Type)

dev_types = data2['DevType'].str.split(';').explode().str.strip()

dev_type_counts = dev_types.value_counts().head(15) # Displaying top 15 for clarity

plt.figure(figsize=(12, 8))

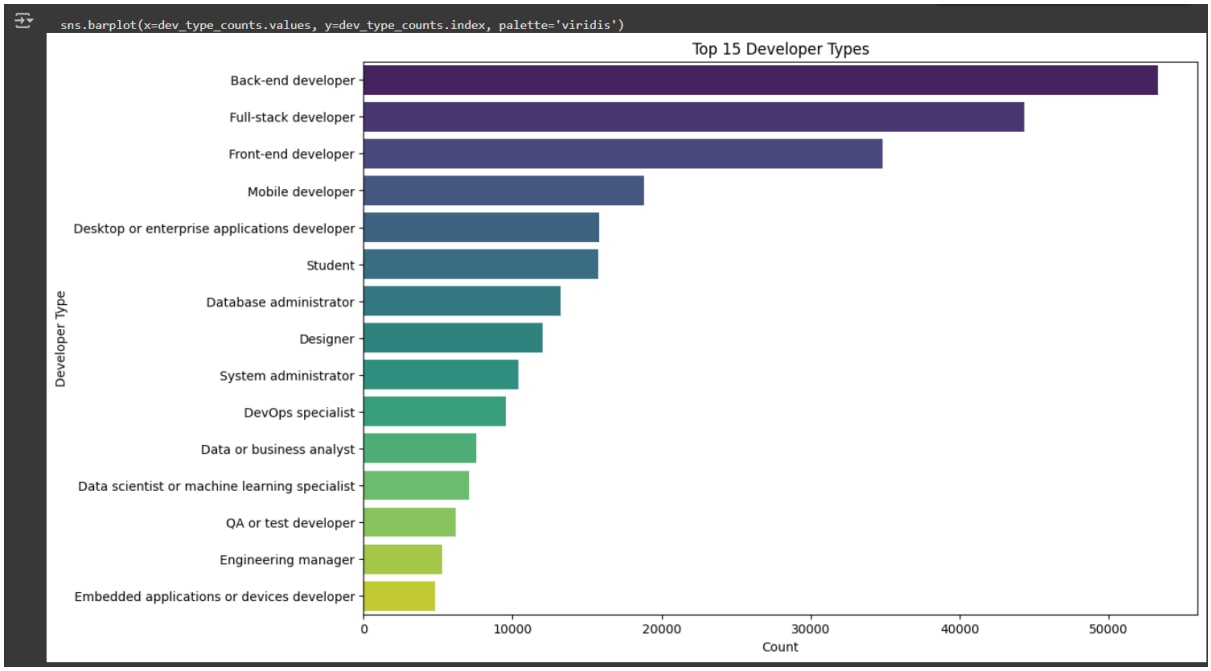
sns.barplot(x=dev_type_counts.values, y=dev_type_counts.index, palette='viridis')

plt.title('Top 15 Developer Types')

plt.xlabel('Count')

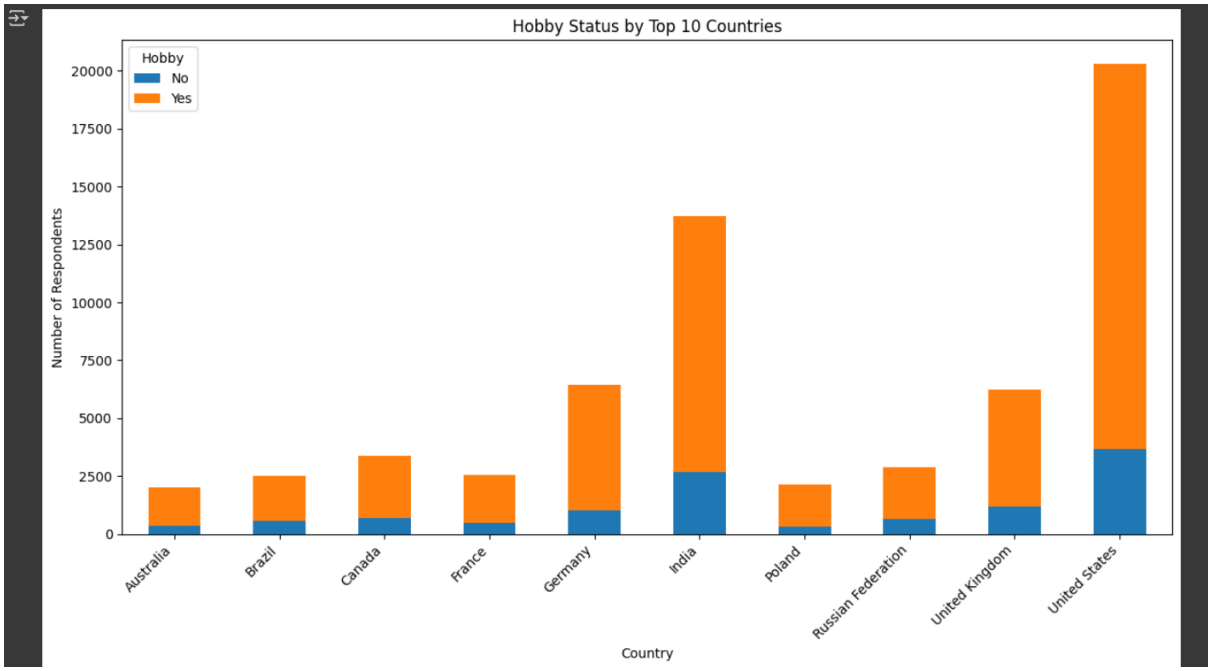
plt.ylabel('Developer Type')

plt.show()



5. Relationship between 'Country' and 'Hobby' for top countries

```
# Get top N countries (e.g., top 10)
top_countries_list = top_20_countries.head(10).index.tolist()
data_top_countries = data2[data2['Country'].isin(top_countries_list)]
hobby_country_counts = pd.crosstab(data_top_countries['Country'], data_top_countries['Hobby'])
hobby_country_counts.plot(kind='bar', stacked=True, figsize=(12, 7))
plt.title('Hobby Status by Top 10 Countries')
plt.xlabel('Country')
plt.ylabel('Number of Respondents')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```


Hobby Status by Top 10 Countries
Hobby
No
Yes
Number of Respondents
0
2500
5000
7500
10000
12500
15000
17500
20000
Australia
Brazil
Canada
France
Germany
India
Poland
Russian Federation
United Kingdom
United States
Country