

INDIAN Start-up Funding Analysis Dashboard

Indian startups have been making waves in recent years, attracting significant attention and funding from investors both domestically & globally.

Aim:

The aim of this project is to analyze funding received by start-ups in India of year (2004, 2008, and 2010 to 2019) and investigate the ecosystem.

Data Collection:

Indian start-up funding dataset of (2004, 2008, 2010 to 2019) year.

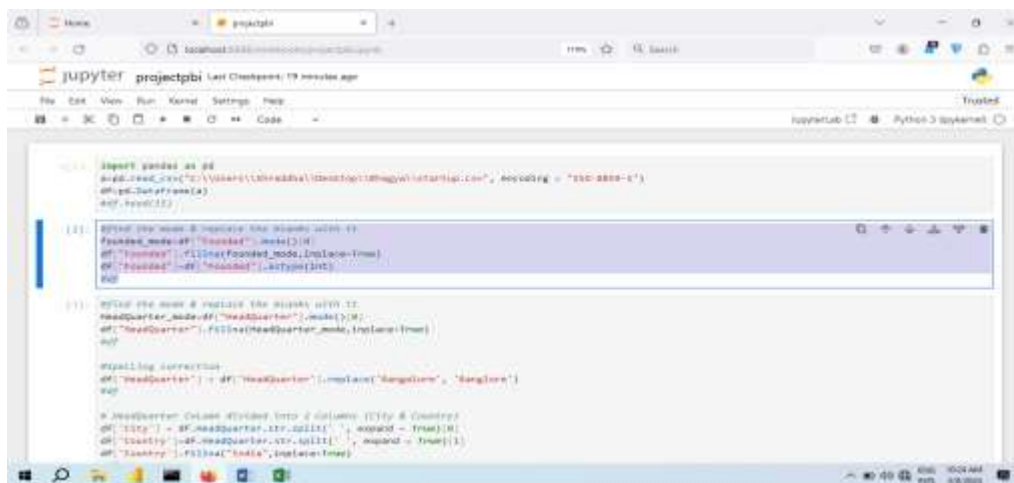
The Column names & Description are as follows:

1. Company/Brand: Name of the Company/Brand/Startup
2. Founded: Year start-up was founded.
3. Sector: Sector of service
4. What it does: Description about Company.
5. Founders: Founders of the Company.
6. Amount (\$): Raised fund.
7. Stage: Round of funding reached.

Data Cleaning:

Cleaning of the data was done with Jupyter Notebook.

1. Founded Column: There were empty cells (Missing values) those were replaced with the mode of the founded column that is most occurring year.



```
import pandas as pd
url = "https://data.cityofchicago.org/api/v1/indian-startup-funding.csv"
df = pd.read_csv(url)

# Step 1: Replace the missing values in the 'Founded' column with the mode
df['Founded'] = df['Founded'].fillna(df['Founded'].mode()[0])

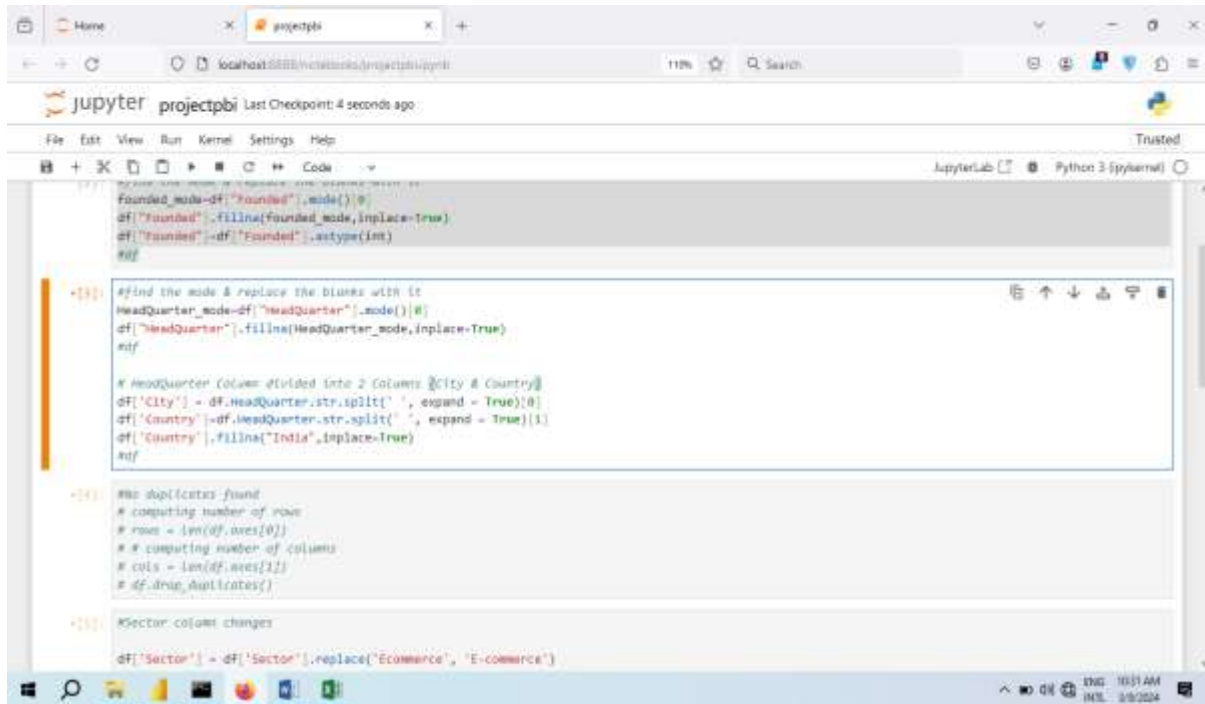
# Step 2: Replace the missing values in the 'HeadQuarter' column with the mode
df['HeadQuarter'] = df['HeadQuarter'].fillna(df['HeadQuarter'].mode()[0])

# Spelling correction
df['HeadQuarter'] = df['HeadQuarter'].replace('Bangalore', 'Banglore')

# Step 3: Split the 'HeadQuarter' column into two columns: 'City' & 'Country'
df['City'] = df['HeadQuarter'].str.split(' ', expand=True)[0]
df['Country'] = df['HeadQuarter'].str.split(' ', expand=True)[1]
df['Country'] = df['Country'].replace('India', 'Indian')
```

2. Headquarter:

- There were empty cells (Missing values) those were replaced with the mode of the founded column that is most occurring Headquarter.
- Headquarter Column divided into 2 Columns (City & Country)



```
[1]: founded_mode=df['Founded'].mode()[0]
df['Founded'].fillna(founded_mode,inplace=True)
df['Founded']=df['Founded'].astype(int)
df

+-----+
[2]: #find the mode & replace the blanks with it
HeadQuarter_mode=df['HeadQuarter'].mode()[0]
df['HeadQuarter'].fillna(HeadQuarter_mode,inplace=True)
df

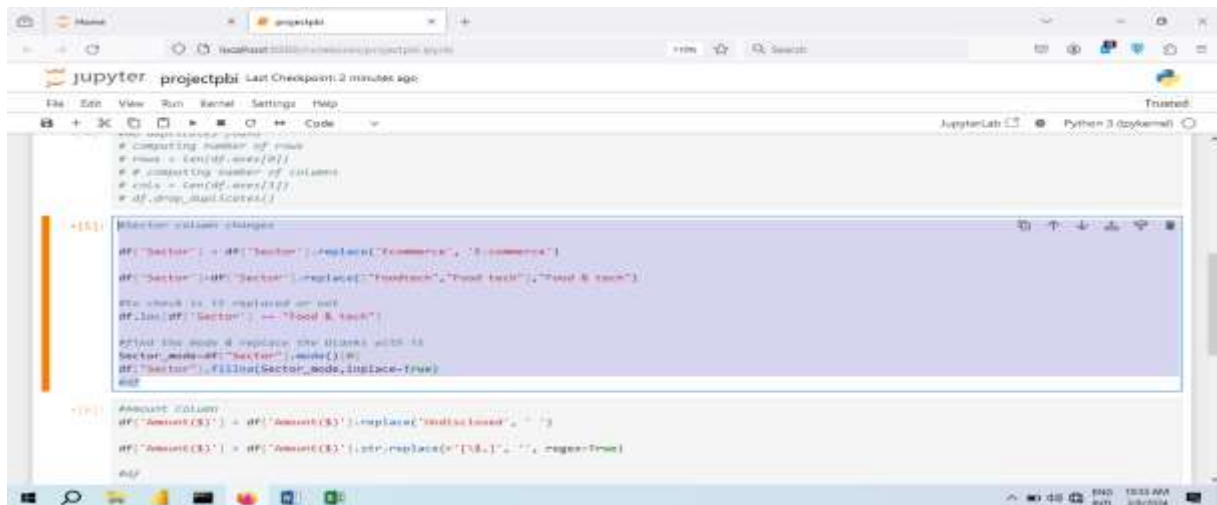
# HeadQuarter column divided into 2 Columns @City & Country@
df['City']=df['HeadQuarter'].str.split(' ',expand=True)[0]
df['Country']=df['HeadQuarter'].str.split(' ',expand=True)[1]
df['Country'].fillna('India',inplace=True)
df

+-----+
[3]: #no duplicates found
# computing number of rows
# rows = len(df.axes[0])
# # computing number of columns
# cols = len(df.axes[1])
# df.drop_duplicates()

+-----+
[4]: #Sector column changes
df['Sector']=df['Sector'].replace('Ecommerce','E-commerce')
```

3. Sector:

- There were empty cells (Missing values) those were replaced with the mode of the founded column that is most occurring Sector.
- There were some cases issues means same names but diff types.so I changed them into one & replaced them.



```
# Computing number of rows
# rows = len(df.axes[0])
# # Computing number of columns
# cols = len(df.axes[1])
# df.drop_duplicates(inplace=True)

#Sector values stages
df['Sector'] = df['Sector'].replace("Commerce", "E-commerce")
df['Sector'] = df['Sector'].replace("Foodtech", "Food tech", "Food & tech")

#to check is it replaced or not
df.loc[df['Sector'] == "Food & tech"]

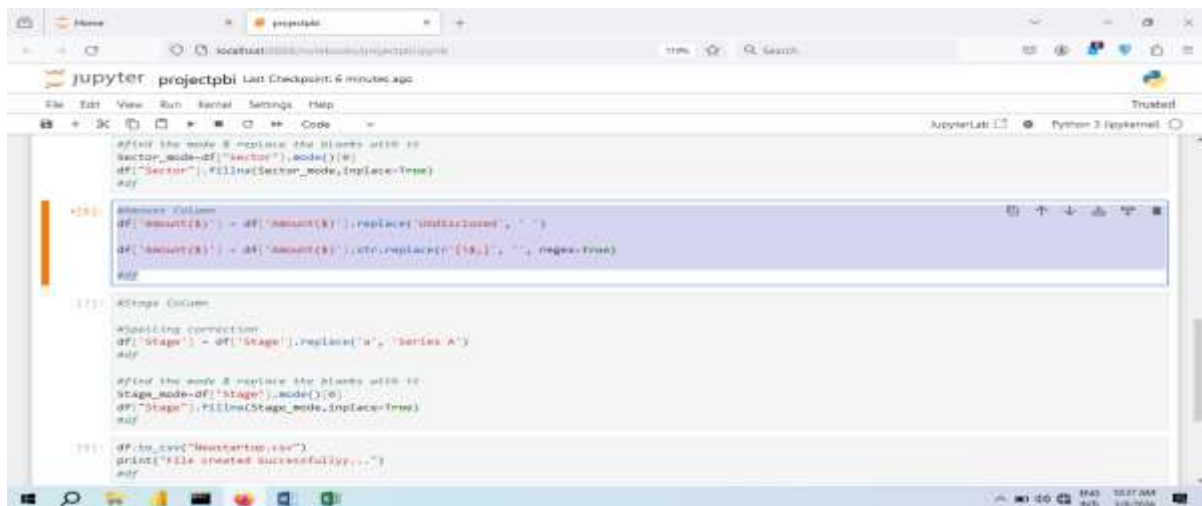
#After the code it replace the blanks with ''
Sector_mode = df['Sector'].mode()[0]
df['Sector'].fillna(Sector_mode, inplace=True)

#Amount Column
df['Amount ($)'] = df['Amount ($)'].replace("undisclosed", "")
df['Amount ($)'] = df['Amount ($)'].str.replace("$", "", regex=True)

df
```

4. Amount:

- The amount column was an integer but had strings of “undisclosed” was replaced with blanks
- Then “\$” and “,” too removed before getting the total values.



```
#After the code it replace the blanks with ''
Sector_mode = df['Sector'].mode()[0]
df['Sector'].fillna(Sector_mode, inplace=True)

#Amount Column
df['Amount ($)'] = df['Amount ($)'].replace("undisclosed", "")
df['Amount ($)'] = df['Amount ($)'].str.replace("$", "", regex=True)

df

#Stage Column
#spelling correction
df['Stage'] = df['Stage'].replace("a", "Series A")

#After the code it replace the blanks with ''
Stage_mode = df['Stage'].mode()[0]
df['Stage'].fillna(Stage_mode, inplace=True)

df.to_csv('newstartup.csv')
print("File created successfully...")

df
```

5. Stage:

- Had similar name & missing values issues. Similar naming replaced with uniform name &
- Missing values replaced with the mode value of the column.

Home projectpbi

localhost:8888/notebooks/projectpbi.ipynb

jupyter projectpbi Last Checkpoint: 9 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code

jupyterlab Python 3 (ipykernel)

```
[1]: Amount Column
df['Amount ($)'] = df['Amount ($)'].replace('Undisclosed', ' ')

df['Amount ($)'] = df['Amount ($)'].str.replace('(\$',)', '', regex=True)
df

[2]: #Stage Column

aligning correction
df['Stage'] = df['Stage'].replace('a', 'Series A')
df

#find the mode & replace the blanks with it
stage_mode=df['Stage'].mode()[0]
df['Stage'].fillna(stage_mode, inplace=True)
df

[3]: df.to_csv("Newstartup.csv")
print("File created Successfully...")
df

File created Successfully...
```

10:40 AM 3/9/2024

Conclusion:

In conclusion,

-The best course of action to take venturing into the Indian start-up ecosystem is venture into an Edtech with its Headquarters in Bangalore.

-Bangalore as shown in the analysis plays a significant role in making startup successful, it has the highest top 5 funded companies.

-Edtech has the highest funded company & also the most popular startup.