

## Case Study - Startup Dataset

This dataset has funding information of the Indian startups from January 2015 to August 2017.

### Feature Details :

**SNo** - Serial number.

**Date** - Date of funding in format DD/MM/YYYY.

**StartupName** - Name of the startup which got funded.

**IndustryVertical** - Industry to which the startup belongs.

**SubVertical** - Sub-category of the industry type.

**CityLocation** - City which the startup is based out of.

**InvestorsName** - Name of the investors involved in the funding round.

**InvestmentType** - Either Private Equity or Seed Funding.

**AmountInUSD** - Funding Amount in USD.

**Remarks** - Other information, if any.

### Insights -

Find out what type of startups are getting funded in the last few years?

Who are the important investors?

What are the hot fields that get a lot of funding these days?

### Questions :-

### Problem Statement :

Check the trend of investments over the years. To check the trend, find -

Total number of fundings done in each year.

Plot a line graph between year and number of fundings. Take year on x-axis and number of fundings on y-axis.

Print year-wise total number of fundings also. Print years in ascending order.

### Note :

There is some error in the 'Date' feature. Make sure to handle that.

### Output Format :

year1 TotalFunding1

year2 TotalFunding2

...

### Problem Statement :

Find out which cities are generally chosen for starting a startup.

Find top 10 Indian cities which have most number of startups ?

Plot a pie chart and visualise it.

Print the city name and number of startups in that city also.

**Note :**

Take city name "Delhi" as "New Delhi".

Check the case-sensitiveness of cities also. That means - at some place, instead of "Bangalore", "bangalore" is given. Take city name as "Bangalore".

For few startups multiple locations are given, one Indian and one Foreign. Count those startups in Indian startup also. Indian city name is first.

Print the city in descending order with respect to the number of startups.

**Problem Statement :**

Find out if cities play any role in receiving funding.

Find top 10 Indian cities with most amount of fundings received. Find out percentage of funding each city has got (among top 10 Indian cities only).

Print the city and percentage with 2 decimal place after rounding off.

**Note:**

Take city name "Delhi" as "New Delhi".

Check the case-sensitiveness of cities also. That means - at some place, instead of "Bangalore", "bangalore" is given. Take city name as "Bangalore".

For few startups multiple locations are given, one Indian and one Foreign. Count those startups in Indian startup also. Indian city name is first.

Print the city in descending order with respect to the percentage of funding.

**Output Format :**

city1 percent1

city2 percent2

city3 percent3

...

...

...

Given File 'startup\_funding.csv'

**Problem Statement :**

There are 4 different type of investments. Find out percentage of amount funded for each investment type.

Plot a pie chart to visualise.

Print the investment type and percentage of amount funded with 2 decimal places after rounding off.

**Note :**

Correct spelling of investment types are - "Private Equity", "Seed Funding", "Debt Funding", and "Crowd Funding". Keep an eye for any spelling mistake. You can find this by printing unique values from this column.

Print the investment type in descending order with respect to the percentage of the amount funded.

**Output Format :**

```
investmentType1 percent1
investmentType2 percent2
investmentType3 percent3
...
```

Given File 'startup\_funding.csv'

**Problem Statement :**

Which type of companies got more easily funding. To answer this question, find -

Top 5 industries and percentage of the total amount funded to that industry. (among top 5 only)

Print the industry name and percentage of the amount funded with 2 decimal place after rounding off.

**Note :**

Ecommerce is the right word in IndustryVertical, so correct it.

Print the industry in descending order with respect to the percentage of the amount funded.

**Output Format :**

```
industry1 percent1
industry2 percent2
industry3 percent3
...
```

**Problem Statement :**

Find top 5 startups with most amount of total funding.

Print the startup name in descending order with respect to amount of funding.

**Note:**

Ola, Flipkart, Oyo, Paytm are important startups, so correct their names. There are many errors in startup names, ignore correcting all, just handle important ones.

**Output Format :**

```
startup1
startup2
startup3
```

```
...
```

**Problem Statement :**

Find the top 5 startups who received the most number of funding rounds. That means, startups which got fundings maximum number of times.

Print the startup name in descending order with respect to the number of funding round as integer value.

**Note:**

Ola, Flipkart, Oyo, Paytm are important startups, so correct their names. There are many errors in startup names, ignore correcting all, just handle important ones.

**Output Format :**

```
startup1 number1
startup2 number2
startup3 number3
```

```
...
```

**DATASET**

[https://drive.google.com/file/d/1YafdMdvwFG3HlwFbVCGiBp3ZJ0EsRlPq/view?usp=drive\\_link](https://drive.google.com/file/d/1YafdMdvwFG3HlwFbVCGiBp3ZJ0EsRlPq/view?usp=drive_link)