

# EDA PROJECT

---

NAME – AMAN GAVEL

REGNO. – I201843I

SECTION – K20RU

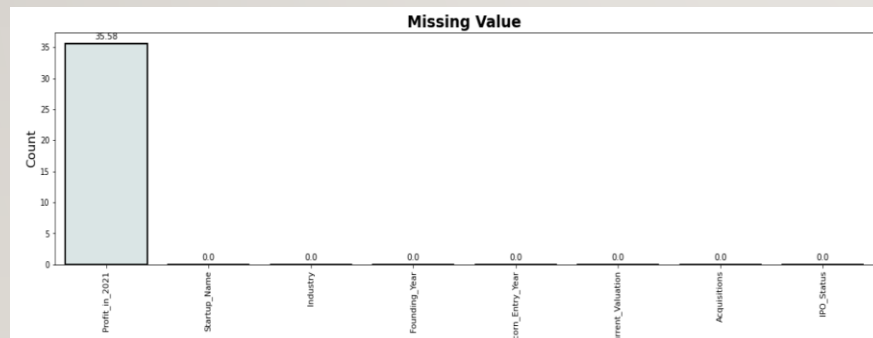


# UNICORN COMPANIES OF INDIA

---

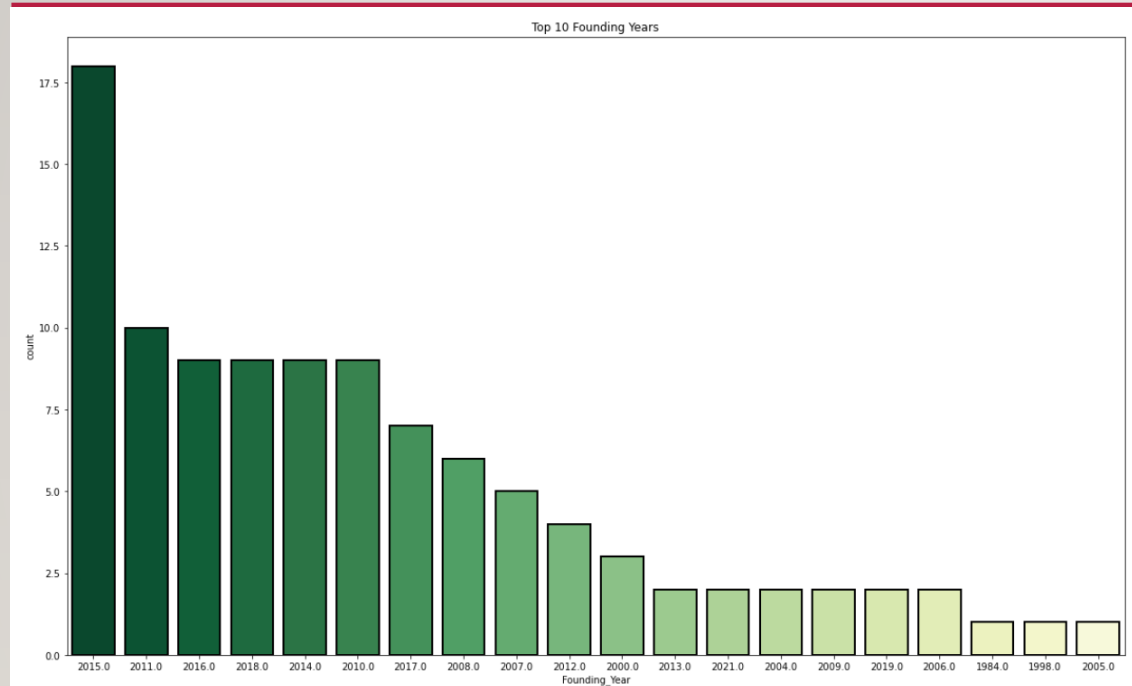
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1858 entries, 0 to 1857
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Startup_Name          929 non-null   object
1   Industry              929 non-null   object
2   Founding_Year         104 non-null   float64
3   Unicorn_Entry_Year    104 non-null   float64
4   Profit_in_2021        67 non-null    object
5   Current_Valuation     104 non-null   object
6   Acquisitions          104 non-null   float64
7   IPO_Status            104 non-null   object
dtypes: float64(3), object(5)
memory usage: 116.2+ KB
```

# GRAPH OF MISSING VALUES



- Here we can see there is a lot of missing value in the dataset.
- Remove the rows which don't have founding year because then we can't analyse them properly
- After removing them we can see that there are very less missing values now we can do Analysis on this.

# TOP 20 FOUNDATION YEARS



- As we can see there are most number of companies formed in 2015 have become unicorn company by 2022.
- Least is from years before 2005
- We can also observe that majority are from the years after the 2010.

THANK YOU

---

