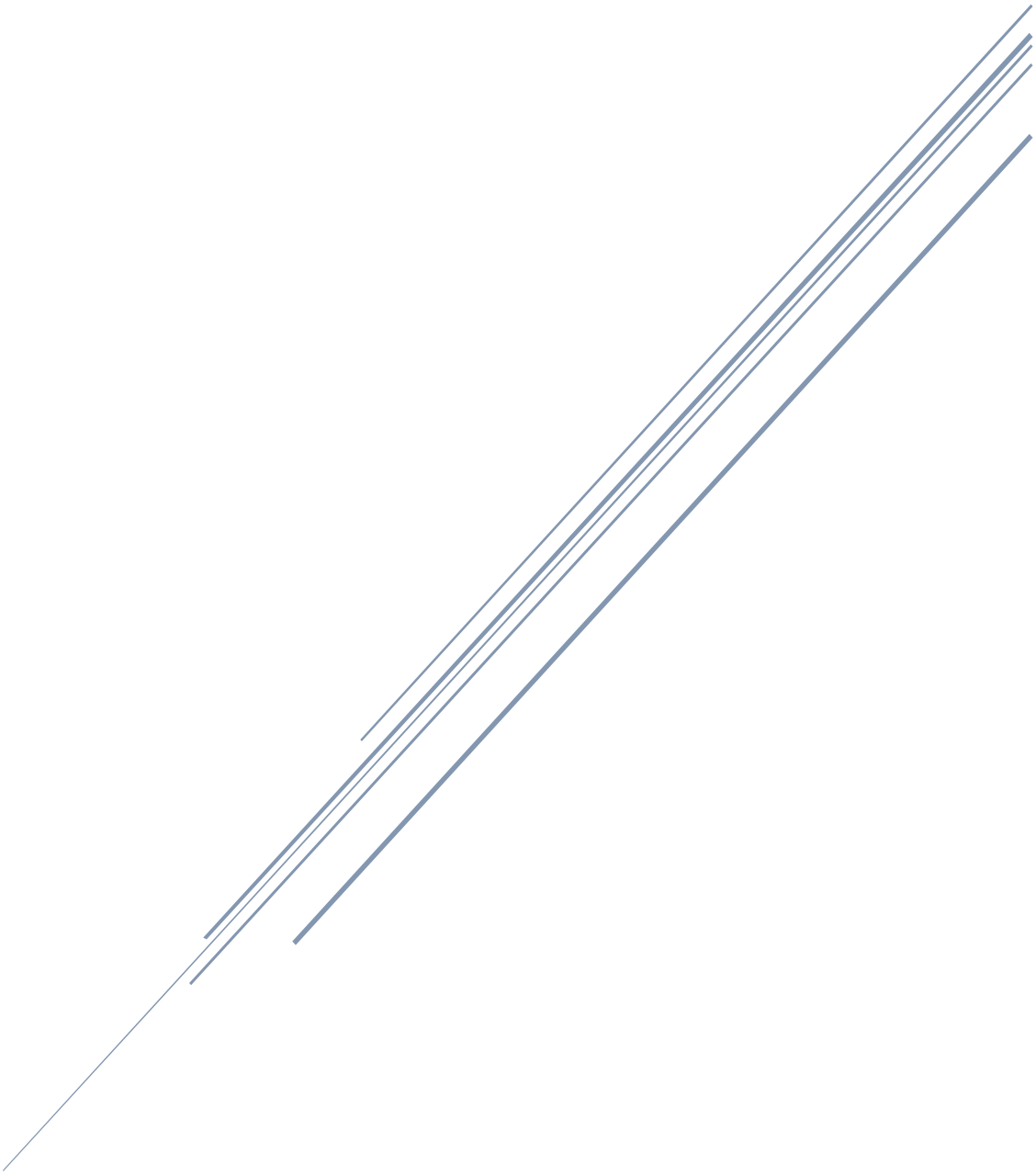
COLLEGEDUNIYA DATA ANALYSIS

By- Yash Wardhan

2020UEE4638



NSUT
2020UEE4638

Overview

Collegedunia Web Pvt Ltd. has successfully established itself as a leading educational portal, India's largest review platform, and an extensive search engine for students, parents, and education industry players who are seeking information on the higher education sector in India and abroad.

Since 2014, Collegedunia has been committed to helping students and parents with authentic information on more than 27,000 colleges, 7000+ courses, 350+ exams segmented in streams like Management, Engineering, Medical, Arts and much more. At Collegdunia, we aim to help students make the right decision, hence offering over 2,00,000 bonafide College Reviews.

Within the span of 5 years, Collegedunia has an average of 10 Lac sessions daily and aims to multiply the number in the coming future. We aspire to reach the maximum number of students who wish to pursue higher education in India or Abroad.

Business Insights

- **Time Frame :**The data spans from '2022-01-04' to'2022-12-08' , covering a period of 338 days.
- **Monthly Distribution :** 23.8% of Leads were generated in July followed by June(19.3%) and August(16.7%).
- **Lead Origin:** 99.5% of the Lead were generated by API and rest 0.5% were through the Online mode.
- **Lead Stage:** 49% of the Lead were Staged as Untouched followed by 21% as Cold Replies.
- **Lead Status:** 78.4% of Lead were from Unverified Sources and rest of them were from verified sources.
- **Course:** BTech had the most number of Enrollment followed by B.Com and MBA .
- **Enrollment Percentage:** 99.84% of the Leads did not lead to enrollment whereas 0.16% of Leads were enrolled in the services of the Company.
- **Campaign Branch:** The campaign branch – 89 yielded the most leads of around 30% followed by 87 (24.6%) and 89(14.7%).

Geographical Focus :

- **State:** Karnataka has the most lead generated of about 32.78% followed by Tamil Nadu(11.64%) and Andra Pradesh(8.94%).
- **Most Leads City:** Bengaluru Rural has the most number of leads generated about 25.92% followed by Chennai 5.30% and Hyderabad 4%.
- **Business Corridor:** Around 40.2% of the business was catered by 5 cities including Bengaluru Rural, Chennai , Hyderabad , Kolkata and New Delhi . Rest 59.8 % of the leads were generated with 1162 cities . Thus, we can infer that Tier-1 cities has a large market for the business

Time Focus:

- **Day of the Week :** Tuesday has the most of the Instance for the Lead created standing at 17.8% ahead of Saturday(15.57%) and Thursday(15.33%).
- **Busiest Hour of the Day:** 12P.M. was the busiest or Last Instance Activity hour in a day standing at 36.9% of Leads last recorded followed by 2P.M. (18.05%) and 6P.M.(14.8%).
- **Leads Activity Duration:** 86.9% of Leads activity lasted for 0 days followed by 0.8% of Lead Activity of 1 day , 0.35% Lead Activity lasted for 2 days.
- **Lead Enrollment Analysis:** From the graphs it is analyzed that it the Lead Activity Duration Lasted between 4-100 days then it has maximum changes of getting lead converted into services.

Business Findings :

Lead Status Distribution

- **Untouched Leads**: The highest number of leads are untouched (74,223). This indicates a significant portion of leads are not being followed up.
- **Cold Leads**: There are 30,147 cold leads, with 57.42% being unverified.
- **Hot Leads**: Only 225 hot leads, with a higher verification rate (59.11%).

Enrolment Status Distribution

Enrolment Status by Course

- Courses with slightly higher enrolment rates include **B. Des** (0.16%) and **BCA** (0.20%).
- **B.Tech.** and **MBA**: Despite a large number of leads, they have low enrolment rates (0.13% and 0.13% respectively).

Enrolment Status by State

- **Higher enrolment rates:** Nagaland (1.24%) followed by Manipur (0.54%) and Himachal Pradesh (0.50%) have a higher enrolment rate. But the number of Leads present in these states are very low (This might be the reason for high percentage).

Business Recommendations

Areas for improvement

Lead Focus

How can we increase the enrolment rate based on the Duration of the Activity Date ?

- From the analysis done , we can infer that the if the duration of lead increases from 2 days to further 100-150 days then the conversion chances of the Lead to Enrolment increases .
- Thus we need to keep the lead in loop and have regular feedback and personalisation of the schemes provided from the company so to keep the customer/lead in better position for enrolment

How can we improve follow-up with leads?

- Automated the follow-up systems to reduce the number of untouched leads as number of untouched leads are more than 49% which might have potential leads of conversion.
- Prioritize follow-up with cold leads to verify and warm them up.

How can we nurture leads more effectively?

- Develop targeted content and campaigns for each stage of the lead journey.
- Personalize communication to increase engagement and conversion rates.

Course Focus

Which courses should we prioritize for marketing efforts?

- Focus marketing and follow-up efforts on courses with higher enrollment rates, like B. Des, BCA, and MBA.
- Reassess courses with 0% enrolment rates, such as B.Sc (Multimedia - SFX, VFX, and Gaming), B.Sc. Economics, BA (Journalism and Mass Communication), BA (Liberal Studies) (Hons), BA - Film and TV Production, and BA Digital Economics (Hons).
- Increase the enrolment rate in high lead courses like Engineering, BCom and MBA by personalising customer college preferences , College predictor and many more features.

Geographical Focus

How can we target our efforts geographically?

- States such as Karnataka , Andra Pradesh and Tamil Nadu have large amount of leads but with low conversion rate . Reassess the customer engagement to convert them into potential leads
- The leads in the States with Karnataka, Andra Pradesh and Tamil Nadu have lead only in top tier cities or the capital city and its nearby region. Try to outreach in more cities with high lead states
- Try to include local language customer support to increase inclusivity and reach within a particular state. Such as Kannada for Karnataka, Tamil for Tamil Nadu , Marathi for Maharashtra etc.
- Analyse reasons behind low enrolment rates in states like Punjab and Goa to improve engagement.

Operational Efficiency

How can we enhance operational efficiency?

- Train the sales and support teams to handle leads more efficiently, especially in high-volume states like Karnataka (47,816 leads) and Uttar Pradesh (11,285 leads).

What role can automation play in improving operations?

- We can use AI-driven chatbots for initial lead interaction and basic queries, particularly for managing the 74,223 untouched leads because we cannot have such a large level of man-power customer support for so many leads

Data Analysis and Feedback

How can we use data to improve our lead management?

- Regularly analyze lead data to identify the areas for improvement.
  For example, study the 8,917 not interested leads to understand their reasons. Also look for feedbacks and trends in the data according to month. Like increase the team potential in July due to college admissions and much more Data Analysis.

- Use predictive analytics to identify high-potential leads and prioritize them. Use Machine Learning Algorithms for those analysis such as Random Forest , Logistic Regression and much more

Implement a feedback loop from unsuccessful leads

- Implement a feedback loop from unsuccessful leads to understand reasons for non-enrolment.
- Make strategies based on feedback to better meet potential students' needs, especially from states with no enrolment like Punjab, Goa, and Chandigarh.

What training should be provided to the sales team?

- Regularly train the sales team on the latest tools and techniques for lead conversion.
- Provide training on soft skills to improve interactions with potential students, focusing on states with high cold lead volumes like Bihar (6,051) and West Bengal (8,913).

Soft Skills Training

What training should be provided to the sales team?

- Regularly train the sales team on the latest tools and techniques for lead conversion.
- Provide training on soft skills to improve interactions with potential students, focusing on states with high cold lead volumes like Bihar (6,051) and West Bengal (8,913).

Data Collection

Which category significantly needs better data collection for lead enrolment optimisation?

- Categories such as Registration Device , Form Initiated , Paid Applications needs data collection and mining . These categories have approx. 99.8 percent of NULL values .
- The increase in data collection of these specific columns can help to better the systems , marketing and many more unseen patterns that can help in increase of enrolment percentage

SUMMARY

This comprehensive analysis of CollegeDuniya's datasets of Geo-Time factors and efficiency reveals significant insights and areas for improvement. By addressing outliers, improving efficiency in approaching leads, and enhancing data collection and quality, CollegeDuniya can optimize its lead generation, increase enrolment conversions, and become a leader in this domain. Implementing the provided recommendations will help achieve better resource (customer support) allocation, improve enrolment percentages, and expand the company's reach, thereby driving operational excellence.

CODE Details :-

Below provided is the Code Snippet followed by:

1)File integration

2)Data Cleaning

3)Feature Engineering

4)Data Visualization

5) Insights through graphs and Pivot table

Note:- I used my sisters laptop for the Purpose of Analysis because my Laptop is currently facing some issues. So, the directory have Yashita written instead of Yash .

```
import pandas as pd
import numpy as np
import matplotlib as plt

file1 = r"C:\Users\Yash\Desktop\College Duniya - Project\Dump1 (1) (1)
(1) (1) (1) (1) (1) (1).xlsx"
file2 = r"C:\Users\Yash\Desktop\College Duniya - Project\Dump2 (1) (1)
(1) (1) (1) (1) (1) (1).xlsx"
file3 = r"C:\Users\Yash\Desktop\College Duniya - Project\Dump3 (1) (1)
(1) (1) (1) (1) (1) (1).xlsx"


df1 = pd.read_excel(file1)
df2 = pd.read_excel(file2)
df3 = pd.read_excel(file3)

# Concatenate the DataFrames
final = pd.concat([df1, df2, df3], ignore_index=True)

# Display the final DataFrame
final.head()
```

```
                  Name                                          Email
Mobile  \
0         Kishan Kumar                       raj******@gmail.com
9986******
1                 LIJO                       lij******@gmail.com
8618******
2     Jeevithachandran             kav****************@gmail.com
8884******
3  CHANDRA shekhar s v  cha**************************@gmail.com
8123******
4        Ravin Kumar G               ram**************@gmail.com
9731******

  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
1         API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
2         API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
3         API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
4         API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022

     Campaign      Lead Stage Lead Status Registration Device  \
0  ABC/89/API            Cold  Unverified                 NaN
```

```
1  ABC/89/API  NOT INTERESTED  Unverified                 NaN
2  ABC/89/API       Untouched  Unverified                 NaN
3  ABC/89/API            Cold  Unverified                 NaN
4  ABC/89/API     No Response  Unverified                 NaN

              Course                 Specialization
Campus  \
0               BCA  Specialization Not Available  University Not
Available
1  B.Sc. Multimedia  Specialization Not Available  University Not
Available
2               BCA  Specialization Not Available  University Not
Available
3               BCA  Specialization Not Available  University Not
Available
4             B.Com  Specialization Not Available  University Not
Available

  Last Lead Activity Date Form Initiated Paid Applications Enrolment
Status
0        06-30-2022 23:55            NaN                NaN
No
1     2022-01-07 10:07:00            NaN                NaN
No
2        06-30-2022 23:50            NaN                NaN
No
3        06-30-2022 23:18            NaN                NaN
No
4        06-30-2022 22:57            NaN                NaN
No
```

```
final.shape
```

```
(150000, 20)
```

```
final.duplicated().sum()
```

```
0
```

```
final.isnull().sum()
```

```
Name                             0
Email                            0
Mobile                           0
Lead Origin                      0
Country                          0
State                            0
City                             0
Instance                         0
Instance Date                    0
Campaign                         0
```

```
Lead Stage                          0
Lead Status                         0
Registration Device            149305
Course                              0
Specialization                      0
Campus                              0
Last Lead Activity Date             0
Form Initiated                 147631
Paid Applications              149396
Enrolment Status                    0
dtype: int64
```

```
final.dtypes
```

```
Name                           object
Email                          object
Mobile                         object
Lead Origin                    object
Country                        object
State                          object
City                           object
Instance                       object
Instance Date                  object
Campaign                       object
Lead Stage                     object
Lead Status                    object
Registration Device            object
Course                         object
Specialization                 object
Campus                         object
Last Lead Activity Date        object
Form Initiated                 object
Paid Applications              object
Enrolment Status               object
dtype: object
```

```
final['Lead Origin'].value_counts()
```

```
API       149305
Online       695
Name: Lead Origin, dtype: int64
```

```
final['Country'].value_counts()
```

```
India                     149992
Bangladesh                     2
United Arab Emirates           2
Oman                           1
Cocos Islands                  1
Saudi Arabia                   1
```

```
Sri Lanka                      1
Name: Country, dtype: int64
```

```
final['State'].value_counts()
```

```
Karnataka                  49171
Tamil Nadu                 17462
Andhra Pradesh             13424
Uttar Pradesh              11460
Kerala                     11276
West Bengal                 9171
Telangana                   7471
Bihar                       6133
Maharashtra                 4178
Madhya Pradesh              3338
Jharkhand                   2804
Rajasthan                   2490
Delhi                       2369
Gujarat                     1711
Odisha                      1401
Assam                       1225
Haryana                      902
Chhattisgarh                 658
Punjab                       456
Uttarakhand                  417
Goa                          411
Jammu and Kashmir            298
Puducherry                   265
Tripura                      208
Himachal Pradesh             204
Arunachal Pradesh            198
Manipur                      189
Nagaland                     164
Andaman and Nicobar          142
Meghalaya                    138
Chandigarh                    98
Sikkim                        75
Mizoram                       56
Dadra And Nagar Haveli        18
Daman And Diu                  8
State Not Available            8
Ladakh                         2
Lakshadweep                    1
Name: State, dtype: int64
```

```
final["Lead Stage"].value_counts()
```

```
Untouched                74285
Cold                     31503
No Response              27020
```

```
NOT INTERESTED            9255
Warm                      3881
Invalid                   2534
Application Submitted      619
NOT ELIGIBLE               581
Hot                        245
Reject                      74
Selected                     3
Name: Lead Stage, dtype: int64
```

```
final["Lead Status"].value_counts()
```

```
Unverified    117747
Verified       32253
Name: Lead Status, dtype: int64
```

```
final["Course"].value_counts()
```

```
B.Tech.                                                          55551
MBA                                                              23647
B.Com                                                            15009
BCA                                                              13566
BBA                                                              10396
LLB                                                               9339
B.Sc Data Science                                                 4234
B.Sc. Economics                                                   4096
B.Sc (Multimedia - SFX, VFX and Gaming)                           4071
B.Sc. Multimedia                                                  4028
B. Des                                                            1966
LL.M.                                                             1960
MA (Psychology)                                                    404
MA (Journalism and Mass Communication)                             404
BA Psychology (Hons)                                               216
BA English (Hons)                                                  178
BA - Film and TV Production                                        162
BA Economics (Hons)                                                160
BA (Journalism and Mass Communication)                             153
BA (Liberal Studies) (Hons)                                        142
BA Digital Economics (Hons)                                        137
BA-Journalism                                                      127
Bachelor of Fine Arts (BFA)                                         39
M.Tech.                                                             14
M.Com - (Forensic Accounting and Corporate Fraud Management)         1
Name: Course, dtype: int64
```

```
final["Specialization"].value_counts()
```

```
Specialization Not Available    150000
Name: Specialization, dtype: int64
```

```
final["Campus"].value_counts()
```

```
University Not Available    150000
Name: Campus, dtype: int64
```

```
final["Form Initiated"].value_counts()
```

```
B.Tech/SOE/Application Form
744
BCA/SOIS/Application Form
312
BBA/SOM/Application Form
293
B.Com/SOC/Application Form
252
MBA/SOM/Application Form
234

...
BSc/SOIS/Application Form,Design Application Form
1
BSC/SOD/Application form,BSc/SOIS/Application Form,International
Admission Form - Applicable only for NRI and Foreign
Nationals,BCA/SOIS/Application Form      1
BA Journalism - School of Media Studies,BSc/SOIS/Application Form,LLB
Form
1
B.Tech/SOE/Application Form,Design Application Form,
BSc/SOC/Application Form,LLB Form
1
BSC/SOD/Application form,B.Tech/SOE/Application Form
1
Name: Form Initiated, Length: 117, dtype: int64
```

```
final["Paid Applications"].value_counts()
```

```
B.Tech/SOE/Application Form
200
BBA/SOM/Application Form
129
BCA/SOIS/Application Form
89
MBA/SOM/Application Form
65
B.Com/SOC/Application Form
57
LLB Form
18
Design Application Form
14
BSc/SOIS/Application Form
5
```

```
LLM/SOL/Application Form
5
 BSc/SOC/Application Form
4
B.Com/SOC/Application Form,BBA/SOM/Application Form
3
BSC/SOD/Application form
2
BBA/SOM/Application Form,BCA/SOIS/Application Form
2
BCA/SOIS/Application Form,B.Tech/SOE/Application Form
2
BA Journalism - School of Media Studies
2
 BSc/SOC/Application Form,BSc/SOIS/Application Form
1
BCA/SOIS/Application Form,BBA/SOM/Application Form
1
International Admission Form - Applicable only for NRI and Foreign
Nationals      1
B.Tech/SOE/Application Form,BSc/SOIS/Application Form
1
B.Tech/SOE/Application Form,Design Application Form
1
Design Application Form,BBA/SOM/Application Form
1
Design Application Form,BCA/SOIS/Application Form
1
Name: Paid Applications, dtype: int64
```

```
final["Enrolment Status"].value_counts()
```

```
No     149762
Yes       238
Name: Enrolment Status, dtype: int64
```

```
final.head(15)
```

```
                    Name                                          Email
Mobile  \
0          Kishan Kumar                        raj******@gmail.com
9986******
1                  LIJO                        lij******@gmail.com
8618******
2      Jeevithachandran              kav***************@gmail.com
8884******
3   CHANDRA shekhar s v  cha*************************@gmail.com
8123******
4         Ravin Kumar G                ram*************@gmail.com
9731******
```

```
5           Afifa Ahmed                    afi*********@gmail.com
9739******
6           Elavarasi M                  ela*********@hotmail.com
9535******
7              BHUVAN P                      bhu******@gmail.com
9731******
8           Jayanth Nag                   jay**********@gmail.com
8884******
9       Milcoln Noronha                  mil***********@gmail.com
7019******
10              Prajwal               the***************@gmail.com
7022******
11    Sananth Kumar M V                  san************@gmail.com
8553******
12            Rachana A                   rac**********@gmail.com
7892******
13          Giri kishan                       gk9*****@gmail.com
9739******
14                Jayson                   jay*********@gmail.com
9900******

   Lead Origin Country      State             City Instance Instance
Date  \
0          API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
1          API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
2          API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
3          API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
4          API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
5          API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
6          API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
7          API   India  Karnataka        Bengaluru  Primary    06-30-
2022
8          API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
9          API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
10         API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
11         API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
12         API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
```

```
13         API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022
14         API   India  Karnataka  Bengaluru Rural  Primary    06-30-
2022

      Campaign      Lead Stage Lead Status Registration Device  \
0   ABC/89/API            Cold  Unverified                 NaN
1   ABC/89/API  NOT INTERESTED  Unverified                 NaN
2   ABC/89/API       Untouched  Unverified                 NaN
3   ABC/89/API            Cold  Unverified                 NaN
4   ABC/89/API     No Response  Unverified                 NaN
5   ABC/89/API            Warm    Verified                 NaN
6   ABC/89/API            Cold  Unverified                 NaN
7   ABC/89/API     No Response    Verified                 NaN
8   ABC/89/API       Untouched  Unverified                 NaN
9   ABC/89/API            Cold    Verified                 NaN
10  ABC/89/API       Untouched  Unverified                 NaN
11  ABC/89/API            Cold    Verified                 NaN
12  ABC/89/API            Cold    Verified                 NaN
13  ABC/89/API     No Response  Unverified                 NaN
14  ABC/89/API            Cold    Verified                 NaN

                                      Course
Specialization  \
0                                        BCA  Specialization Not
Available
1                           B.Sc. Multimedia  Specialization Not
Available
2                                        BCA  Specialization Not
Available
3                                        BCA  Specialization Not
Available
4                                      B.Com  Specialization Not
Available
5                                        BBA  Specialization Not
Available
6                           B.Sc. Multimedia  Specialization Not
Available
7   B.Sc (Multimedia - SFX, VFX and Gaming)  Specialization Not
Available
8                            B.Sc. Economics  Specialization Not
Available
9                                    B. Des  Specialization Not
Available
10                                  B.Tech.  Specialization Not
Available
11                                       BCA  Specialization Not
Available
12                                     LL.M.  Specialization Not
```

```
Available
13                                        BCA  Specialization Not
Available
14                                        BCA  Specialization Not
Available

                      Campus Last Lead Activity Date  \
0   University Not Available        06-30-2022 23:55
1   University Not Available     2022-01-07 10:07:00
2   University Not Available        06-30-2022 23:50
3   University Not Available        06-30-2022 23:18
4   University Not Available        06-30-2022 22:57
5   University Not Available     2022-04-07 04:06:00
6   University Not Available        06-30-2022 22:46
7   University Not Available        07-14-2022 12:17
8   University Not Available        06-30-2022 22:46
9   University Not Available        06-30-2022 22:41
10  University Not Available        06-30-2022 22:41
11  University Not Available     2022-01-07 23:52:00
12  University Not Available        07-13-2022 19:20
13  University Not Available        06-30-2022 22:20
14  University Not Available        06-30-2022 22:20

                Form Initiated Paid Applications Enrolment Status
0                          NaN                NaN               No
1                          NaN                NaN               No
2                          NaN                NaN               No
3                          NaN                NaN               No
4                          NaN                NaN               No
5                          NaN                NaN               No
6                          NaN                NaN               No
7   BSc/SOIS/Application Form                 NaN               No
8                          NaN                NaN               No
9                          NaN                NaN               No
10                         NaN                NaN               No
11                         NaN                NaN               No
12                         NaN                NaN               No
13                         NaN                NaN               No
14                         NaN                NaN               No
```

```
final["Campaign"].value_counts()
```

```
ABC/86/API              45071
ABC/87/API              37027
ABC/89/API              22145
ABC/88/API              21756
ABC/79/API               8443
ABC/86_RV/API            7330
ABC/70/API               2135
ABC/90_EI/API            1309
```

```
ABC/84/API                1031
ABC/88_RV/API              894
ABC/90/API                 713
ABC/91/API                 595
ABC/87_RV/API              555
ABC/Liveform/Online        332
ABC/Emailer/Online         293
ABC/84_RV/API              292
ABC/SMS/Online              44
ABC/Emailer/NA              20
ABC/MLink/MLink              6
ABC/91_RV/API                5
ABC/90_RV/API                4
Name: Campaign, dtype: int64
```

```
final.tail()
```

```
                    Name                            Email      Mobile  \
149995   Rahul senapati       sra**********@gmail.com  8695******
149996        Rohit dey         roh*********@gmail.com  8777******
149997            K SAI   aka**************@gmail.com  8085******
149998  Swapnil Samanta    sam**************@gmail.com  7063******
149999     Shruti Gupta  shr****************@gmail.com  7003******

       Lead Origin Country        State               City Instance  \
149995         API   India  West Bengal  South 24 Parganas  Primary
149996         API   India  West Bengal            Kalyani  Primary
149997         API   India  West Bengal            Kolkata  Primary
149998         API   India  West Bengal            Kolkata  Primary
149999         API   India  West Bengal            Kolkata  Primary

       Instance Date    Campaign      Lead Stage Lead Status  \
149995    07-16-2022  ABC/87/API       Untouched  Unverified
149996    07-16-2022  ABC/87/API       Untouched  Unverified
149997    07-16-2022  ABC/87/API       Untouched  Unverified
149998    07-16-2022  ABC/87/API       Untouched  Unverified
149999    07-16-2022  ABC/87/API  NOT INTERESTED    Verified

       Registration Device   Course               Specialization  \
149995                 NaN      BBA  Specialization Not Available
149996                 NaN      BBA  Specialization Not Available
149997                 NaN  B.Tech.  Specialization Not Available
149998                 NaN  B.Tech.  Specialization Not Available
149999                 NaN      BBA  Specialization Not Available

                         Campus Last Lead Activity Date Form
Initiated  \
149995  University Not Available        07-16-2022 12:29
NaN
149996  University Not Available        07-16-2022 12:29
```

```
NaN
149997  University Not Available        07-16-2022 12:29
NaN
149998  University Not Available        07-16-2022 12:29
NaN
149999  University Not Available     2022-08-09 13:06:00
NaN

       Paid Applications Enrolment Status
149995               NaN               No
149996               NaN               No
149997               NaN               No
149998               NaN               No
149999               NaN               No
```

```
final["City"].value_counts()
```

```
Bengaluru Rural    38888
Chennai             7963
Hyderabad           6162
Kolkata             5758
New Delhi           2363
                   ...
Tenga Market           1
Mudargi                1
Kanhangad              1
Yadgir                 1
Dapoli                 1
Name: City, Length: 1167, dtype: int64
```

```
final.describe()
```

```
            Name                         Email      Mobile Lead Origin
Country  \
count     150000                        150000      150000      150000
150000
unique    118891                         35662        2801           2
7
top     Abhishek  sha************@gmail.com  9353******         API
India
freq         147                           366         731      149305
149992

            State             City Instance Instance Date    Campaign
\
count      150000           150000   150000        150000      150000

unique         38             1167        1           163          21

top     Karnataka  Bengaluru Rural  Primary    07-22-2022  ABC/86/API
```

```
freq        49171            38888   150000          2923       45071


       Lead Stage Lead Status Registration Device   Course  \
count      150000      150000                 695   150000
unique         11           2                   3       25
top     Untouched  Unverified              Mobile  B.Tech.
freq        74285      117747                 583    55551

                      Specialization                    Campus  \
count                         150000                    150000
unique                             1                         1
top     Specialization Not Available  University Not Available
freq                          150000                    150000

       Last Lead Activity Date               Form Initiated  \
count                   150000                         2369
unique                   26845                          117
top           04-21-2022 16:02  B.Tech/SOE/Application Form
freq                       284                          744

                  Paid Applications Enrolment Status
count                           604           150000
unique                           22                2
top     B.Tech/SOE/Application Form               No
freq                            200           149762
```

```
final["Instance"].value_counts()
```

```
Primary    150000
Name: Instance, dtype: int64
```

```
final["Form Initiated"].value_counts()
```

```
B.Tech/SOE/Application Form
744
BCA/SOIS/Application Form
312
BBA/SOM/Application Form
293
B.Com/SOC/Application Form
252
MBA/SOM/Application Form
234

...
BSc/SOIS/Application Form,Design Application Form
1
BSC/SOD/Application form,BSc/SOIS/Application Form,International
Admission Form - Applicable only for NRI and Foreign
Nationals,BCA/SOIS/Application Form      1
```

```
BA Journalism - School of Media Studies,BSc/SOIS/Application Form,LLB
Form
1
B.Tech/SOE/Application Form,Design Application Form,
BSc/SOC/Application Form,LLB Form
1
BSC/SOD/Application form,B.Tech/SOE/Application Form
1
Name: Form Initiated, Length: 117, dtype: int64
```

```
final["Paid Applications"].value_counts()
```

```
B.Tech/SOE/Application Form
200
BBA/SOM/Application Form
129
BCA/SOIS/Application Form
89
MBA/SOM/Application Form
65
B.Com/SOC/Application Form
57
LLB Form
18
Design Application Form
14
BSc/SOIS/Application Form
5
LLM/SOL/Application Form
5
 BSc/SOC/Application Form
4
B.Com/SOC/Application Form,BBA/SOM/Application Form
3
BSC/SOD/Application form
2
BBA/SOM/Application Form,BCA/SOIS/Application Form
2
BCA/SOIS/Application Form,B.Tech/SOE/Application Form
2
BA Journalism - School of Media Studies
2
 BSc/SOC/Application Form,BSc/SOIS/Application Form
1
BCA/SOIS/Application Form,BBA/SOM/Application Form
1
International Admission Form - Applicable only for NRI and Foreign
Nationals      1
B.Tech/SOE/Application Form,BSc/SOIS/Application Form
1
```

```
B.Tech/SOE/Application Form,Design Application Form
1
Design Application Form,BBA/SOM/Application Form
1
Design Application Form,BCA/SOIS/Application Form
1
Name: Paid Applications, dtype: int64
```

```
final["Enrolment Status"].value_counts()
```

```
No     149762
Yes       238
Name: Enrolment Status, dtype: int64
```

```
final.duplicated().sum()
```

```
0
```

```
final.isnull().sum()
```

```
Name                           0
Email                          0
Mobile                         0
Lead Origin                    0
Country                        0
State                          0
City                           0
Instance                       0
Instance Date                  0
Campaign                       0
Lead Stage                     0
Lead Status                    0
Registration Device        149305
Course                         0
Specialization                 0
Campus                         0
Last Lead Activity Date        0
Form Initiated             147631
Paid Applications          149396
Enrolment Status               0
dtype: int64
```

```
final.dtypes
```

```
Name                       object
Email                      object
Mobile                     object
Lead Origin                object
Country                    object
State                      object
City                       object
```

```
Instance                    object
Instance Date               object
Campaign                    object
Lead Stage                  object
Lead Status                 object
Registration Device         object
Course                      object
Specialization              object
Campus                      object
Last Lead Activity Date     object
Form Initiated              object
Paid Applications           object
Enrolment Status            object
dtype: object
```

```
final["Name"].dtypes
```

```
dtype('O')
```

```
final["Instance Date"] = pd.to_datetime(final["Instance Date"])
final["Last Lead Activity Date"]= pd.to_datetime(final["Last Lead
Activity Date"])
```

```
final.dtypes
```

```
Name                                object
Email                               object
Mobile                              object
Lead Origin                         object
Country                             object
State                               object
City                                object
Instance                            object
Instance Date               datetime64[ns]
Campaign                            object
Lead Stage                          object
Lead Status                         object
Registration Device                 object
Course                              object
Specialization                      object
Campus                              object
Last Lead Activity Date     datetime64[ns]
Form Initiated                      object
Paid Applications                   object
Enrolment Status                    object
dtype: object
```

```
final.head()
```

```
                  Name                                           Email
Mobile  \
```

```
0         Kishan Kumar                          raj******@gmail.com
9986******
1                 LIJO                          lij******@gmail.com
8618******
2     Jeevithachandran              kav****************@gmail.com
8884******
3  CHANDRA shekhar s v  cha**************************@gmail.com
8123******
4        Ravin Kumar G                ram**************@gmail.com
9731******

  Lead Origin Country      State              City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
1         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

     Campaign      Lead Stage Lead Status Registration Device  \
0  ABC/89/API            Cold  Unverified                 NaN
1  ABC/89/API  NOT INTERESTED  Unverified                 NaN
2  ABC/89/API       Untouched  Unverified                 NaN
3  ABC/89/API            Cold  Unverified                 NaN
4  ABC/89/API     No Response  Unverified                 NaN

             Course                 Specialization
Campus  \
0               BCA  Specialization Not Available  University Not
Available
1  B.Sc. Multimedia  Specialization Not Available  University Not
Available
2               BCA  Specialization Not Available  University Not
Available
3               BCA  Specialization Not Available  University Not
Available
4             B.Com  Specialization Not Available  University Not
Available

  Last Lead Activity Date Form Initiated Paid Applications Enrolment
Status
0     2022-06-30 23:55:00            NaN               NaN
No
1     2022-01-07 10:07:00            NaN               NaN
No
```

```
2     2022-06-30 23:50:00             NaN                  NaN
No
3     2022-06-30 23:18:00             NaN                  NaN
No
4     2022-06-30 22:57:00             NaN                  NaN
No
```

```
final["Specialization"].value_counts()
```

```
Specialization Not Available    150000
Name: Specialization, dtype: int64
```

```
final.drop(columns=["Specialization"])
```

```
                        Name                                      Email
\
0               Kishan Kumar                        raj******@gmail.com

1                       LIJO                        lij******@gmail.com

2          Jeevithachandran                kav***************@gmail.com

3       CHANDRA shekhar s v  cha**************************@gmail.com

4              Ravin Kumar G                ram*************@gmail.com

...                      ...                                       ...

149995       Rahul senapati                   sra**********@gmail.com

149996            Rohit dey                     roh*********@gmail.com

149997                K SAI                aka**************@gmail.com

149998      Swapnil Samanta                sam*************@gmail.com

149999         Shruti Gupta               shr***************@gmail.com


            Mobile Lead Origin Country        State                City
\
0       9986******         API   India    Karnataka    Bengaluru Rural

1       8618******         API   India    Karnataka    Bengaluru Rural

2       8884******         API   India    Karnataka    Bengaluru Rural

3       8123******         API   India    Karnataka    Bengaluru Rural

4       9731******         API   India    Karnataka    Bengaluru Rural

...            ...         ...     ...          ...                ...
```

```
149995  8695******         API   India  West Bengal  South 24 Parganas
149996  8777******         API   India  West Bengal            Kalyani
149997  8085******         API   India  West Bengal            Kolkata
149998  7063******         API   India  West Bengal            Kolkata
149999  7003******         API   India  West Bengal            Kolkata

       Instance Instance Date    Campaign      Lead Stage Lead Status
\
0       Primary    2022-06-30  ABC/89/API            Cold  Unverified
1       Primary    2022-06-30  ABC/89/API  NOT INTERESTED  Unverified
2       Primary    2022-06-30  ABC/89/API       Untouched  Unverified
3       Primary    2022-06-30  ABC/89/API            Cold  Unverified
4       Primary    2022-06-30  ABC/89/API     No Response  Unverified
...         ...           ...         ...             ...         ...
149995  Primary    2022-07-16  ABC/87/API       Untouched  Unverified
149996  Primary    2022-07-16  ABC/87/API       Untouched  Unverified
149997  Primary    2022-07-16  ABC/87/API       Untouched  Unverified
149998  Primary    2022-07-16  ABC/87/API       Untouched  Unverified
149999  Primary    2022-07-16  ABC/87/API  NOT INTERESTED    Verified

       Registration Device            Course                     Campus
\
0                     NaN               BCA  University Not Available
1                     NaN  B.Sc. Multimedia  University Not Available
2                     NaN               BCA  University Not Available
3                     NaN               BCA  University Not Available
4                     NaN             B.Com  University Not Available
...                   ...               ...                       ...
149995                NaN               BBA  University Not Available
```

```
149996                  NaN                BBA  University Not Available

149997                  NaN           B.Tech.  University Not Available

149998                  NaN           B.Tech.  University Not Available

149999                  NaN                BBA  University Not Available


       Last Lead Activity Date Form Initiated Paid Applications  \
0          2022-06-30 23:55:00            NaN                NaN
1          2022-01-07 10:07:00            NaN                NaN
2          2022-06-30 23:50:00            NaN                NaN
3          2022-06-30 23:18:00            NaN                NaN
4          2022-06-30 22:57:00            NaN                NaN
...                        ...            ...                ...
149995     2022-07-16 12:29:00            NaN                NaN
149996     2022-07-16 12:29:00            NaN                NaN
149997     2022-07-16 12:29:00            NaN                NaN
149998     2022-07-16 12:29:00            NaN                NaN
149999     2022-08-09 13:06:00            NaN                NaN

       Enrolment Status
0                    No
1                    No
2                    No
3                    No
4                    No
...                 ...
149995               No
149996               No
149997               No
149998               No
149999               No

[150000 rows x 19 columns]

final.drop(columns=["Name","Email","Mobile"])

       Lead Origin Country        State               City Instance  \
0               API   India    Karnataka    Bengaluru Rural  Primary
1               API   India    Karnataka    Bengaluru Rural  Primary
2               API   India    Karnataka    Bengaluru Rural  Primary
3               API   India    Karnataka    Bengaluru Rural  Primary
4               API   India    Karnataka    Bengaluru Rural  Primary
...             ...     ...          ...                ...      ...
149995          API   India  West Bengal  South 24 Parganas  Primary
149996          API   India  West Bengal            Kalyani  Primary
149997          API   India  West Bengal            Kolkata  Primary
```

```
149998         API   India  West Bengal            Kolkata  Primary
149999         API   India  West Bengal            Kolkata  Primary

       Instance Date    Campaign      Lead Stage Lead Status  \
0         2022-06-30  ABC/89/API            Cold  Unverified
1         2022-06-30  ABC/89/API  NOT INTERESTED  Unverified
2         2022-06-30  ABC/89/API       Untouched  Unverified
3         2022-06-30  ABC/89/API            Cold  Unverified
4         2022-06-30  ABC/89/API     No Response  Unverified
...              ...         ...             ...         ...
149995    2022-07-16  ABC/87/API       Untouched  Unverified
149996    2022-07-16  ABC/87/API       Untouched  Unverified
149997    2022-07-16  ABC/87/API       Untouched  Unverified
149998    2022-07-16  ABC/87/API       Untouched  Unverified
149999    2022-07-16  ABC/87/API  NOT INTERESTED    Verified

       Registration Device            Course
Specialization  \
0                       NaN               BCA  Specialization Not
Available
1                       NaN  B.Sc. Multimedia  Specialization Not
Available
2                       NaN               BCA  Specialization Not
Available
3                       NaN               BCA  Specialization Not
Available
4                       NaN             B.Com  Specialization Not
Available
...                     ...               ...
...
149995                  NaN               BBA  Specialization Not
Available
149996                  NaN               BBA  Specialization Not
Available
149997                  NaN           B.Tech.  Specialization Not
Available
149998                  NaN           B.Tech.  Specialization Not
Available
149999                  NaN               BBA  Specialization Not
Available

                          Campus Last Lead Activity Date Form
Initiated  \
0       University Not Available     2022-06-30 23:55:00
NaN
1       University Not Available     2022-01-07 10:07:00
NaN
2       University Not Available     2022-06-30 23:50:00
NaN
3       University Not Available     2022-06-30 23:18:00
```

```
NaN
4       University Not Available     2022-06-30 22:57:00
NaN
...                           ...                     ...             ..
.
149995  University Not Available     2022-07-16 12:29:00
NaN
149996  University Not Available     2022-07-16 12:29:00
NaN
149997  University Not Available     2022-07-16 12:29:00
NaN
149998  University Not Available     2022-07-16 12:29:00
NaN
149999  University Not Available     2022-08-09 13:06:00
NaN

       Paid Applications Enrolment Status
0                    NaN               No
1                    NaN               No
2                    NaN               No
3                    NaN               No
4                    NaN               No
...                  ...              ...
149995               NaN               No
149996               NaN               No
149997               NaN               No
149998               NaN               No
149999               NaN               No

[150000 rows x 17 columns]
```

```
final.head()
```

```
                  Name                                        Email
Mobile  \
0         Kishan Kumar                       raj******@gmail.com
9986******
1                 LIJO                       lij******@gmail.com
8618******
2     Jeevithachandran             kav***************@gmail.com
8884******
3  CHANDRA shekhar s v  cha**************************@gmail.com
8123******
4        Ravin Kumar G               ram*************@gmail.com
9731******

  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
```

```
1         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

     Campaign      Lead Stage Lead Status Registration Device  \
0  ABC/89/API            Cold  Unverified                NaN
1  ABC/89/API  NOT INTERESTED  Unverified                NaN
2  ABC/89/API       Untouched  Unverified                NaN
3  ABC/89/API            Cold  Unverified                NaN
4  ABC/89/API     No Response  Unverified                NaN

              Course                 Specialization
Campus  \
0               BCA  Specialization Not Available  University Not
Available
1  B.Sc. Multimedia  Specialization Not Available  University Not
Available
2               BCA  Specialization Not Available  University Not
Available
3               BCA  Specialization Not Available  University Not
Available
4             B.Com  Specialization Not Available  University Not
Available

  Last Lead Activity Date Form Initiated Paid Applications Enrolment
Status
0     2022-06-30 23:55:00            NaN               NaN
No
1     2022-01-07 10:07:00            NaN               NaN
No
2     2022-06-30 23:50:00            NaN               NaN
No
3     2022-06-30 23:18:00            NaN               NaN
No
4     2022-06-30 22:57:00            NaN               NaN
No
```

```
final.drop(columns=["Name", "Email", "Specialization","Mobile"],
axis=1, inplace=True)

final.head()
```

```
  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
```

```
30
1         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

     Campaign      Lead Stage Lead Status Registration Device  \
0  ABC/89/API            Cold  Unverified                 NaN
1  ABC/89/API  NOT INTERESTED  Unverified                 NaN
2  ABC/89/API       Untouched  Unverified                 NaN
3  ABC/89/API            Cold  Unverified                 NaN
4  ABC/89/API     No Response  Unverified                 NaN

              Course                     Campus Last Lead Activity Date
\
0               BCA  University Not Available     2022-06-30 23:55:00

1  B.Sc. Multimedia  University Not Available     2022-01-07 10:07:00

2               BCA  University Not Available     2022-06-30 23:50:00

3               BCA  University Not Available     2022-06-30 23:18:00

4              B.Com  University Not Available     2022-06-30 22:57:00


  Form Initiated Paid Applications Enrolment Status
0            NaN               NaN               No
1            NaN               NaN               No
2            NaN               NaN               No
3            NaN               NaN               No
4            NaN               NaN               No
```

```
final["Registration Device"].value_counts()
```

```
Mobile     583
Desktop    110
Tablet       2
Name: Registration Device, dtype: int64
```

```
final["Registration Device"].isnull().sum()
```

```
149305
```

```
final["Country"].isnull().sum()
```

```
0
```

```
arr = final.columns
arr
```

```
Index(['Lead Origin', 'Country', 'State', 'City', 'Instance',
'Instance Date',
       'Campaign', 'Lead Stage', 'Lead Status', 'Registration Device',
       'Course', 'Campus', 'Last Lead Activity Date', 'Form
Initiated',
       'Paid Applications', 'Enrolment Status'],
      dtype='object')
```

```
for i in arr:
    print(final[arr].isnull().sum())
```

```
Lead Origin                     0
Country                         0
State                           0
City                            0
Instance                        0
Instance Date                   0
Campaign                        0
Lead Stage                      0
Lead Status                     0
Registration Device        149305
Course                          0
Campus                          0
Last Lead Activity Date         0
Form Initiated             147631
Paid Applications          149396
Enrolment Status                0
dtype: int64
Lead Origin                     0
Country                         0
State                           0
City                            0
Instance                        0
Instance Date                   0
Campaign                        0
Lead Stage                      0
Lead Status                     0
Registration Device        149305
Course                          0
Campus                          0
Last Lead Activity Date         0
Form Initiated             147631
Paid Applications          149396
Enrolment Status                0
dtype: int64
Lead Origin                     0
Country                         0
```

```
State                              0
City                               0
Instance                           0
Instance Date                      0
Campaign                           0
Lead Stage                         0
Lead Status                        0
Registration Device           149305
Course                             0
Campus                             0
Last Lead Activity Date            0
Form Initiated                147631
Paid Applications             149396
Enrolment Status                   0
dtype: int64
Lead Origin                        0
Country                            0
State                              0
City                               0
Instance                           0
Instance Date                      0
Campaign                           0
Lead Stage                         0
Lead Status                        0
Registration Device           149305
Course                             0
Campus                             0
Last Lead Activity Date            0
Form Initiated                147631
Paid Applications             149396
Enrolment Status                   0
dtype: int64
Lead Origin                        0
Country                            0
State                              0
City                               0
Instance                           0
Instance Date                      0
Campaign                           0
Lead Stage                         0
Lead Status                        0
Registration Device           149305
Course                             0
Campus                             0
Last Lead Activity Date            0
Form Initiated                147631
Paid Applications             149396
Enrolment Status                   0
dtype: int64
```

```
Lead Origin                       0
Country                           0
State                             0
City                              0
Instance                          0
Instance Date                     0
Campaign                          0
Lead Stage                        0
Lead Status                       0
Registration Device          149305
Course                            0
Campus                            0
Last Lead Activity Date           0
Form Initiated               147631
Paid Applications            149396
Enrolment Status                  0
dtype: int64
Lead Origin                       0
Country                           0
State                             0
City                              0
Instance                          0
Instance Date                     0
Campaign                          0
Lead Stage                        0
Lead Status                       0
Registration Device          149305
Course                            0
Campus                            0
Last Lead Activity Date           0
Form Initiated               147631
Paid Applications            149396
Enrolment Status                  0
dtype: int64
Lead Origin                       0
Country                           0
State                             0
City                              0
Instance                          0
Instance Date                     0
Campaign                          0
Lead Stage                        0
Lead Status                       0
Registration Device          149305
Course                            0
Campus                            0
Last Lead Activity Date           0
Form Initiated               147631
Paid Applications            149396
```

```
Enrolment Status                0
dtype: int64
Lead Origin                     0
Country                         0
State                           0
City                            0
Instance                        0
Instance Date                   0
Campaign                        0
Lead Stage                      0
Lead Status                     0
Registration Device        149305
Course                          0
Campus                          0
Last Lead Activity Date         0
Form Initiated             147631
Paid Applications          149396
Enrolment Status                0
dtype: int64
Lead Origin                     0
Country                         0
State                           0
City                            0
Instance                        0
Instance Date                   0
Campaign                        0
Lead Stage                      0
Lead Status                     0
Registration Device        149305
Course                          0
Campus                          0
Last Lead Activity Date         0
Form Initiated             147631
Paid Applications          149396
Enrolment Status                0
dtype: int64
Lead Origin                     0
Country                         0
State                           0
City                            0
Instance                        0
Instance Date                   0
Campaign                        0
Lead Stage                      0
Lead Status                     0
Registration Device        149305
Course                          0
Campus                          0
Last Lead Activity Date         0
```

```
Form Initiated             147631
Paid Applications          149396
Enrolment Status                0
dtype: int64
Lead Origin                     0
Country                         0
State                           0
City                            0
Instance                        0
Instance Date                   0
Campaign                        0
Lead Stage                      0
Lead Status                     0
Registration Device        149305
Course                          0
Campus                          0
Last Lead Activity Date         0
Form Initiated             147631
Paid Applications          149396
Enrolment Status                0
dtype: int64
Lead Origin                     0
Country                         0
State                           0
City                            0
Instance                        0
Instance Date                   0
Campaign                        0
Lead Stage                      0
Lead Status                     0
Registration Device        149305
Course                          0
Campus                          0
Last Lead Activity Date         0
Form Initiated             147631
Paid Applications          149396
Enrolment Status                0
dtype: int64
Lead Origin                     0
Country                         0
State                           0
City                            0
Instance                        0
Instance Date                   0
Campaign                        0
Lead Stage                      0
Lead Status                     0
Registration Device        149305
Course                          0
```

```
Campus                            0
Last Lead Activity Date           0
Form Initiated               147631
Paid Applications            149396
Enrolment Status                  0
dtype: int64
Lead Origin                       0
Country                           0
State                             0
City                              0
Instance                          0
Instance Date                     0
Campaign                          0
Lead Stage                        0
Lead Status                       0
Registration Device          149305
Course                            0
Campus                            0
Last Lead Activity Date           0
Form Initiated               147631
Paid Applications            149396
Enrolment Status                  0
dtype: int64

Lead Origin                       0
Country                           0
State                             0
City                              0
Instance                          0
Instance Date                     0
Campaign                          0
Lead Stage                        0
Lead Status                       0
Registration Device          149305
Course                            0
Campus                            0
Last Lead Activity Date           0
Form Initiated               147631
Paid Applications            149396
Enrolment Status                  0
dtype: int64

final.shape

(150000, 16)

final.dtypes

Lead Origin                      object
Country                          object
```

```
State                                  object
City                                   object
Instance                               object
Instance Date                  datetime64[ns]
Campaign                               object
Lead Stage                             object
Lead Status                            object
Registration Device                    object
Course                                 object
Campus                                 object
Last Lead Activity Date        datetime64[ns]
Form Initiated                         object
Paid Applications                      object
Enrolment Status                       object
dtype: object
```

```
final.head()
```

```
  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
1         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

     Campaign      Lead Stage Lead Status Registration Device  \
0  ABC/89/API            Cold  Unverified                 NaN
1  ABC/89/API  NOT INTERESTED  Unverified                 NaN
2  ABC/89/API       Untouched  Unverified                 NaN
3  ABC/89/API            Cold  Unverified                 NaN
4  ABC/89/API     No Response  Unverified                 NaN

              Course                     Campus Last Lead Activity Date
\
0               BCA  University Not Available     2022-06-30 23:55:00

1  B.Sc. Multimedia  University Not Available     2022-01-07 10:07:00

2               BCA  University Not Available     2022-06-30 23:50:00

3               BCA  University Not Available     2022-06-30 23:18:00

4             B.Com  University Not Available     2022-06-30 22:57:00
```

```
  Form Initiated Paid Applications Enrolment Status
0            NaN               NaN                No
1            NaN               NaN                No
2            NaN               NaN                No
3            NaN               NaN                No
4            NaN               NaN                No
```

```
final["Lead Origin"].value_counts()
```

```
API       149305
Online       695
Name: Lead Origin, dtype: int64
```

```
final["Instance"].value_counts()
```

```
Primary    150000
Name: Instance, dtype: int64
```

```
final["Instance Date"].value_counts()
```

```
2022-07-22    2923
2022-04-08    2838
2022-07-29    2608
2022-07-20    2604
2022-08-17    2540
              ...
2022-09-03      15
2022-03-14      15
2022-08-03      11
2022-05-04       7
2022-12-02       1
Name: Instance Date, Length: 163, dtype: int64
```

```
final["Lead Stage"].value_counts()
```

```
Untouched                74285
Cold                     31503
No Response              27020
NOT INTERESTED            9255
Warm                      3881
Invalid                   2534
Application Submitted      619
NOT ELIGIBLE               581
Hot                        245
Reject                      74
Selected                     3
Name: Lead Stage, dtype: int64
```

```
final["Lead Status"].value_counts()
```

```
Unverified    117747
Verified       32253
Name: Lead Status, dtype: int64
```

```
final["Registration Device"].value_counts()
```

```
Mobile     583
Desktop    110
Tablet       2
Name: Registration Device, dtype: int64
```

```
final["Campaign"].value_counts()
```

```
ABC/86/API              45071
ABC/87/API              37027
ABC/89/API              22145
ABC/88/API              21756
ABC/79/API               8443
ABC/86_RV/API            7330
ABC/70/API               2135
ABC/90_EI/API            1309
ABC/84/API               1031
ABC/88_RV/API             894
ABC/90/API                713
ABC/91/API                595
ABC/87_RV/API             555
ABC/Liveform/Online       332
ABC/Emailer/Online        293
ABC/84_RV/API             292
ABC/SMS/Online             44
ABC/Emailer/NA             20
ABC/MLink/MLink             6
ABC/91_RV/API               5
ABC/90_RV/API               4
Name: Campaign, dtype: int64
```

```
final["Lead Status"].value_counts()
```

```
Unverified    117747
Verified       32253
Name: Lead Status, dtype: int64
```

```
final["Registration Device"].value_counts()
```

```
Mobile     583
Desktop    110
Tablet       2
Name: Registration Device, dtype: int64
```

```
final["Course"].value_counts()
```

```
B.Tech.                                                                55551
MBA                                                                    23647
B.Com                                                                  15009
BCA                                                                    13566
BBA                                                                    10396
LLB                                                                     9339
B.Sc Data Science                                                       4234
B.Sc. Economics                                                         4096
B.Sc (Multimedia - SFX, VFX and Gaming)                                 4071
B.Sc. Multimedia                                                        4028
B. Des                                                                  1966
LL.M.                                                                   1960
MA (Psychology)                                                          404
MA (Journalism and Mass Communication)                                   404
BA Psychology (Hons)                                                     216
BA English (Hons)                                                        178
BA - Film and TV Production                                              162
BA Economics (Hons)                                                      160
BA (Journalism and Mass Communication)                                   153
BA (Liberal Studies) (Hons)                                              142
BA Digital Economics (Hons)                                              137
BA-Journalism                                                            127
Bachelor of Fine Arts (BFA)                                               39
M.Tech.                                                                   14
M.Com - (Forensic Accounting and Corporate Fraud Management)               1
Name: Course, dtype: int64
```

```
final["Campus"].value_counts()
```

```
University Not Available    150000
Name: Campus, dtype: int64
```

```
final["Form Initiated"].value_counts()
```

```
B.Tech/SOE/Application Form
744
BCA/SOIS/Application Form
312
BBA/SOM/Application Form
293
B.Com/SOC/Application Form
252
MBA/SOM/Application Form
234

...
BSc/SOIS/Application Form,Design Application Form
1
BSC/SOD/Application form,BSc/SOIS/Application Form,International
Admission Form - Applicable only for NRI and Foreign
```

```
Nationals,BCA/SOIS/Application Form      1
BA Journalism - School of Media Studies,BSc/SOIS/Application Form,LLB
Form
1
B.Tech/SOE/Application Form,Design Application Form,
BSc/SOC/Application Form,LLB Form
1
BSC/SOD/Application form,B.Tech/SOE/Application Form
1
Name: Form Initiated, Length: 117, dtype: int64
```

```
final["Paid Applications"].value_counts()
```

```
B.Tech/SOE/Application Form
200
BBA/SOM/Application Form
129
BCA/SOIS/Application Form
89
MBA/SOM/Application Form
65
B.Com/SOC/Application Form
57
LLB Form
18
Design Application Form
14
BSc/SOIS/Application Form
5
LLM/SOL/Application Form
5
 BSc/SOC/Application Form
4
B.Com/SOC/Application Form,BBA/SOM/Application Form
3
BSC/SOD/Application form
2
BBA/SOM/Application Form,BCA/SOIS/Application Form
2
BCA/SOIS/Application Form,B.Tech/SOE/Application Form
2
BA Journalism - School of Media Studies
2
 BSc/SOC/Application Form,BSc/SOIS/Application Form
1
BCA/SOIS/Application Form,BBA/SOM/Application Form
1
International Admission Form - Applicable only for NRI and Foreign
Nationals      1
B.Tech/SOE/Application Form,BSc/SOIS/Application Form
```

```
1
B.Tech/SOE/Application Form,Design Application Form
1
Design Application Form,BBA/SOM/Application Form
1
Design Application Form,BCA/SOIS/Application Form
1
Name: Paid Applications, dtype: int64
```

```
final["Enrolment Status"].value_counts()
```

```
No     149762
Yes       238
Name: Enrolment Status, dtype: int64
```

```
final.fillna({'Registration Device': 'Unknown', 'Form Initiated': 'Not
Initiated', 'Paid Applications': 'Not Paid'}, inplace=True)
```

```
final.head()
```

```
  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
1         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

     Campaign      Lead Stage Lead Status Registration Device  \
0  ABC/89/API            Cold  Unverified             Unknown
1  ABC/89/API  NOT INTERESTED  Unverified             Unknown
2  ABC/89/API       Untouched  Unverified             Unknown
3  ABC/89/API            Cold  Unverified             Unknown
4  ABC/89/API     No Response  Unverified             Unknown

              Course                    Campus Last Lead Activity Date
\
0               BCA  University Not Available     2022-06-30 23:55:00

1  B.Sc. Multimedia  University Not Available     2022-01-07 10:07:00

2               BCA  University Not Available     2022-06-30 23:50:00

3               BCA  University Not Available     2022-06-30 23:18:00

4             B.Com  University Not Available     2022-06-30 22:57:00
```

```
  Form Initiated Paid Applications Enrolment Status
0  Not Initiated          Not Paid               No
1  Not Initiated          Not Paid               No
2  Not Initiated          Not Paid               No
3  Not Initiated          Not Paid               No
4  Not Initiated          Not Paid               No
```

```
data=final
```

```
split_columns = data['Campaign'].str.split('/', expand=True)
```

```
split_columns
```

```
          0   1    2
0       ABC  89  API
1       ABC  89  API
2       ABC  89  API
3       ABC  89  API
4       ABC  89  API
...     ...  ..  ...
149995  ABC  87  API
149996  ABC  87  API
149997  ABC  87  API
149998  ABC  87  API
149999  ABC  87  API

[150000 rows x 3 columns]
```

```
data["Organisation"] = split_columns[0]
data["Campaign Number"] = split_columns[1]
data["Channel"] = split_columns[2]
```

```
data.head()
```

```
  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
1         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

     Campaign      Lead Stage Lead Status Registration Device  \
0  ABC/89/API            Cold  Unverified             Unknown
```

```
1  ABC/89/API  NOT INTERESTED  Unverified             Unknown
2  ABC/89/API       Untouched  Unverified             Unknown
3  ABC/89/API            Cold  Unverified             Unknown
4  ABC/89/API     No Response  Unverified             Unknown

              Course                     Campus Last Lead Activity Date
\
0               BCA  University Not Available     2022-06-30 23:55:00

1  B.Sc. Multimedia  University Not Available     2022-01-07 10:07:00

2               BCA  University Not Available     2022-06-30 23:50:00

3               BCA  University Not Available     2022-06-30 23:18:00

4             B.Com  University Not Available     2022-06-30 22:57:00


  Form Initiated Paid Applications Enrolment Status Organisation  \
0  Not Initiated          Not Paid               No          ABC
1  Not Initiated          Not Paid               No          ABC
2  Not Initiated          Not Paid               No          ABC
3  Not Initiated          Not Paid               No          ABC
4  Not Initiated          Not Paid               No          ABC

  Campaign Number Channel
0              89     API
1              89     API
2              89     API
3              89     API
4              89     API
```

```
data.drop(columns=["Campaign"], axis=1 , inplace=True)
```

```
data.head()
```

```
  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
1         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

       Lead Stage Lead Status Registration Device            Course  \
0            Cold  Unverified             Unknown               BCA
```

```
1  NOT INTERESTED  Unverified             Unknown  B.Sc. Multimedia
2       Untouched  Unverified             Unknown               BCA
3            Cold  Unverified             Unknown               BCA
4     No Response  Unverified             Unknown             B.Com

                     Campus Last Lead Activity Date Form Initiated  \
0  University Not Available     2022-06-30 23:55:00  Not Initiated
1  University Not Available     2022-01-07 10:07:00  Not Initiated
2  University Not Available     2022-06-30 23:50:00  Not Initiated
3  University Not Available     2022-06-30 23:18:00  Not Initiated
4  University Not Available     2022-06-30 22:57:00  Not Initiated

  Paid Applications Enrolment Status Organisation Campaign Number
Channel
0          Not Paid                No          ABC               89
API
1          Not Paid                No          ABC               89
API
2          Not Paid                No          ABC               89
API
3          Not Paid                No          ABC               89
API
4          Not Paid                No          ABC               89
API

data.dtypes

Lead Origin                          object
Country                              object
State                                object
City                                 object
Instance                             object
Instance Date                datetime64[ns]
Lead Stage                           object
Lead Status                          object
Registration Device                  object
Course                               object
Campus                               object
Last Lead Activity Date      datetime64[ns]
Form Initiated                       object
Paid Applications                    object
Enrolment Status                     object
Organisation                         object
Campaign Number                      object
Channel                              object
dtype: object

data.dtypes
```

```
Lead Origin                          object
Country                              object
State                                object
City                                 object
Instance                             object
Instance Date              datetime64[ns]
Lead Stage                           object
Lead Status                          object
Registration Device                  object
Course                               object
Campus                               object
Last Lead Activity Date    datetime64[ns]
Form Initiated                       object
Paid Applications                    object
Enrolment Status                     object
Organisation                         object
Campaign Number                      object
Channel                              object
dtype: object
```

```
data["City"].value_counts()
```

```
Bengaluru Rural    38888
Chennai             7963
Hyderabad           6162
Kolkata             5758
New Delhi           2363
                   ...
Tenga Market           1
Mudargi                1
Kanhangad              1
Yadgir                 1
Dapoli                 1
Name: City, Length: 1167, dtype: int64
```

```
data['Days to Last Activity'] = (data['Last Lead Activity Date'] -
data['Instance Date']).dt.days
```

```
data.head()
```

```
  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
1         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
```

```
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

       Lead Stage Lead Status Registration Device            Course  \
0            Cold  Unverified             Unknown               BCA
1  NOT INTERESTED  Unverified             Unknown  B.Sc. Multimedia
2       Untouched  Unverified             Unknown               BCA
3            Cold  Unverified             Unknown               BCA
4     No Response  Unverified             Unknown              B.Com

                     Campus Last Lead Activity Date Form Initiated  \
0  University Not Available     2022-06-30 23:55:00  Not Initiated
1  University Not Available     2022-01-07 10:07:00  Not Initiated
2  University Not Available     2022-06-30 23:50:00  Not Initiated
3  University Not Available     2022-06-30 23:18:00  Not Initiated
4  University Not Available     2022-06-30 22:57:00  Not Initiated

  Paid Applications Enrolment Status Organisation Campaign Number
Channel  \
0          Not Paid               No          ABC               89
API
1          Not Paid               No          ABC               89
API
2          Not Paid               No          ABC               89
API
3          Not Paid               No          ABC               89
API
4          Not Paid               No          ABC               89
API

   Days to Last Activity
0                      0
1                   -174
2                      0
3                      0
4                      0
```

```
data["Days to Last Activity"].value_counts()
```

```
 0      130484
 1        1298
 2         538
 31        387
 3         346
         ...
-256         1
-123         1
-156         1
-224         1
```

```
-208         1
Name: Days to Last Activity, Length: 562, dtype: int64

data.head()

  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
1         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

       Lead Stage Lead Status Registration Device            Course  \
0            Cold  Unverified             Unknown               BCA
1  NOT INTERESTED  Unverified             Unknown  B.Sc. Multimedia
2       Untouched  Unverified             Unknown               BCA
3            Cold  Unverified             Unknown               BCA
4     No Response  Unverified             Unknown             B.Com

                      Campus Last Lead Activity Date Form Initiated  \
0  University Not Available     2022-06-30 23:55:00  Not Initiated
1  University Not Available     2022-01-07 10:07:00  Not Initiated
2  University Not Available     2022-06-30 23:50:00  Not Initiated
3  University Not Available     2022-06-30 23:18:00  Not Initiated
4  University Not Available     2022-06-30 22:57:00  Not Initiated

  Paid Applications Enrolment Status Organisation Campaign Number
Channel  \
0          Not Paid               No          ABC               89
API
1          Not Paid               No          ABC               89
API
2          Not Paid               No          ABC               89
API
3          Not Paid               No          ABC               89
API
4          Not Paid               No          ABC               89
API

   Days to Last Activity
0                      0
1                   -174
2                      0
```

```
3                       0
4                       0
```

```
print(df1.shape)
```

```
(50000, 20)
```

```
print(df2.shape)
```

```
(50000, 20)
```

```
print(df3.shape)
```

```
(50000, 20)
```

```
data['Campaign Number'].value_counts()
```

```
86          45071
87          37027
89          22145
88          21756
79           8443
86_RV        7330
70           2135
90_EI        1309
84           1031
88_RV         894
90            713
91            595
87_RV         555
Liveform      332
Emailer       313
84_RV         292
SMS            44
MLink           6
91_RV           5
90_RV           4
Name: Campaign Number, dtype: int64
```

```
data.dtypes
```

```
Lead Origin                          object
Country                              object
State                                object
City                                 object
Instance                             object
Instance Date              datetime64[ns]
Lead Stage                           object
Lead Status                          object
Registration Device                  object
Course                               object
```

```
Campus                              object
Last Lead Activity Date    datetime64[ns]
Form Initiated                      object
Paid Applications                   object
Enrolment Status                    object
Organisation                        object
Campaign Number                     object
Channel                             object
Days to Last Activity                int64
dtype: object
```

```
negative_days = data[data['Days to Last Activity'] < 0]
print(f"Number of negative entries: {len(negative_days)}")
```

```
Number of negative entries: 3474
```

```
data = data[data['Days to Last Activity'] >= 0]
```

```
data.shape
```

```
(146526, 19)
```

```
data.head()
```

```
  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
6         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

    Lead Stage Lead Status Registration Device            Course  \
0         Cold  Unverified             Unknown               BCA
2    Untouched  Unverified             Unknown               BCA
3         Cold  Unverified             Unknown               BCA
4  No Response  Unverified             Unknown             B.Com
6         Cold  Unverified             Unknown  B.Sc. Multimedia

                     Campus Last Lead Activity Date Form Initiated  \
0  University Not Available     2022-06-30 23:55:00  Not Initiated
2  University Not Available     2022-06-30 23:50:00  Not Initiated
3  University Not Available     2022-06-30 23:18:00  Not Initiated
4  University Not Available     2022-06-30 22:57:00  Not Initiated
6  University Not Available     2022-06-30 22:46:00  Not Initiated
```

```
  Paid Applications Enrolment Status Organisation Campaign Number
Channel  \
0          Not Paid                No          ABC                89
API
2          Not Paid                No          ABC                89
API
3          Not Paid                No          ABC                89
API
4          Not Paid                No          ABC                89
API
6          Not Paid                No          ABC                89
API

   Days to Last Activity
0                      0
2                      0
3                      0
4                      0
6                      0
```

```
data['weekday'] = data['Instance Date'].dt.day_name()
data['day'] = data['Instance Date'].dt.day
data['hour'] = data['Last Lead Activity Date'].dt.hour
data['minute'] = data['Last Lead Activity Date'].dt.minute
```

```
C:\Users\Yashita\AppData\Local\Temp\ipykernel_14064\3922378014.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  data['weekday'] = data['Instance Date'].dt.day_name()
C:\Users\Yashita\AppData\Local\Temp\ipykernel_14064\3922378014.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  data['day'] = data['Instance Date'].dt.day
C:\Users\Yashita\AppData\Local\Temp\ipykernel_14064\3922378014.py:3:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
```

```
returning-a-view-versus-a-copy
  data['hour'] = data['Last Lead Activity Date'].dt.hour
C:\Users\Yashita\AppData\Local\Temp\ipykernel_14064\3922378014.py:4:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  data['minute'] = data['Last Lead Activity Date'].dt.minute

data.head()

  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
6         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

    Lead Stage Lead Status Registration Device            Course  ...
\
0         Cold  Unverified             Unknown               BCA  ...

2    Untouched  Unverified             Unknown               BCA  ...

3         Cold  Unverified             Unknown               BCA  ...

4  No Response  Unverified             Unknown             B.Com  ...

6         Cold  Unverified             Unknown  B.Sc. Multimedia  ...


  Paid Applications Enrolment Status Organisation Campaign Number
Channel  \
0          Not Paid               No          ABC              89
API
2          Not Paid               No          ABC              89
API
3          Not Paid               No          ABC              89
API
4          Not Paid               No          ABC              89
API
```

```
6          Not Paid                No          ABC                89
API

  Days to Last Activity   weekday day  hour minute
0                     0  Thursday  30    23     55
2                     0  Thursday  30    23     50
3                     0  Thursday  30    23     18
4                     0  Thursday  30    22     57
6                     0  Thursday  30    22     46

[5 rows x 23 columns]
```

```
data["Lead Status"].value_counts()
```

```
Unverified    116997
Verified       29529
Name: Lead Status, dtype: int64
```

```
data["month"]=data['Instance Date'].dt.month_name()
```

```
C:\Users\Yashita\AppData\Local\Temp\ipykernel_14064\2101951569.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  data["month"]=data['Instance Date'].dt.month_name()
```

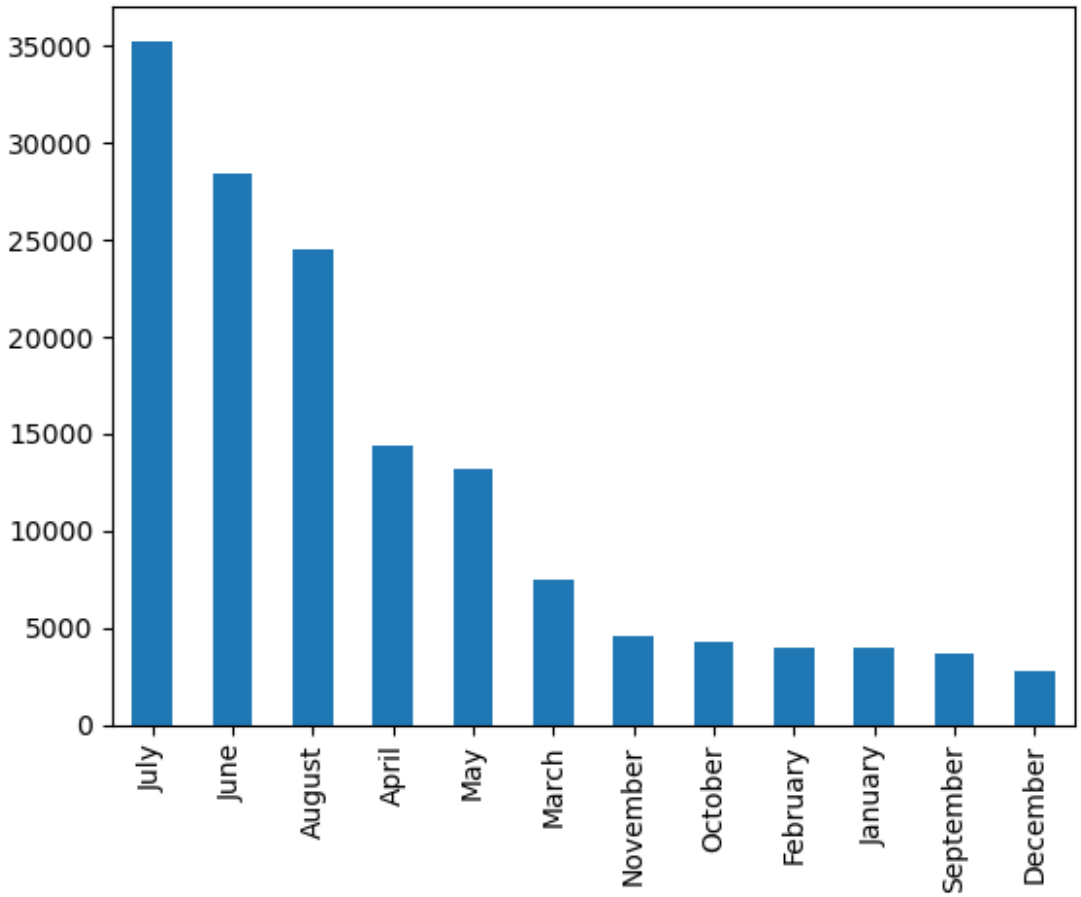
```
data.head()
```

```
  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
6         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

    Lead Stage Lead Status Registration Device            Course  ...
\
0         Cold  Unverified             Unknown               BCA  ...

2    Untouched  Unverified             Unknown               BCA  ...
```

```
3         Cold  Unverified              Unknown                BCA  ...

4  No Response  Unverified              Unknown              B.Com  ...

6         Cold  Unverified              Unknown  B.Sc. Multimedia  ...


  Enrolment Status Organisation Campaign Number Channel Days to Last
Activity  \
0               No          ABC               89     API
0
2               No          ABC               89     API
0
3               No          ABC               89     API
0
4               No          ABC               89     API
0
6               No          ABC               89     API
0

    weekday day hour  minute month
0  Thursday  30   23      55  June
2  Thursday  30   23      50  June
3  Thursday  30   23      18  June
4  Thursday  30   22      57  June
6  Thursday  30   22      46  June

[5 rows x 24 columns]
```

Leads generated v/s the Month

```
data['month'].value_counts().plot(kind='bar')

<Axes: >
```



So we can infer that the Maximum number of Leads generated were in the Month of July followed by June then August and as follows in the graph

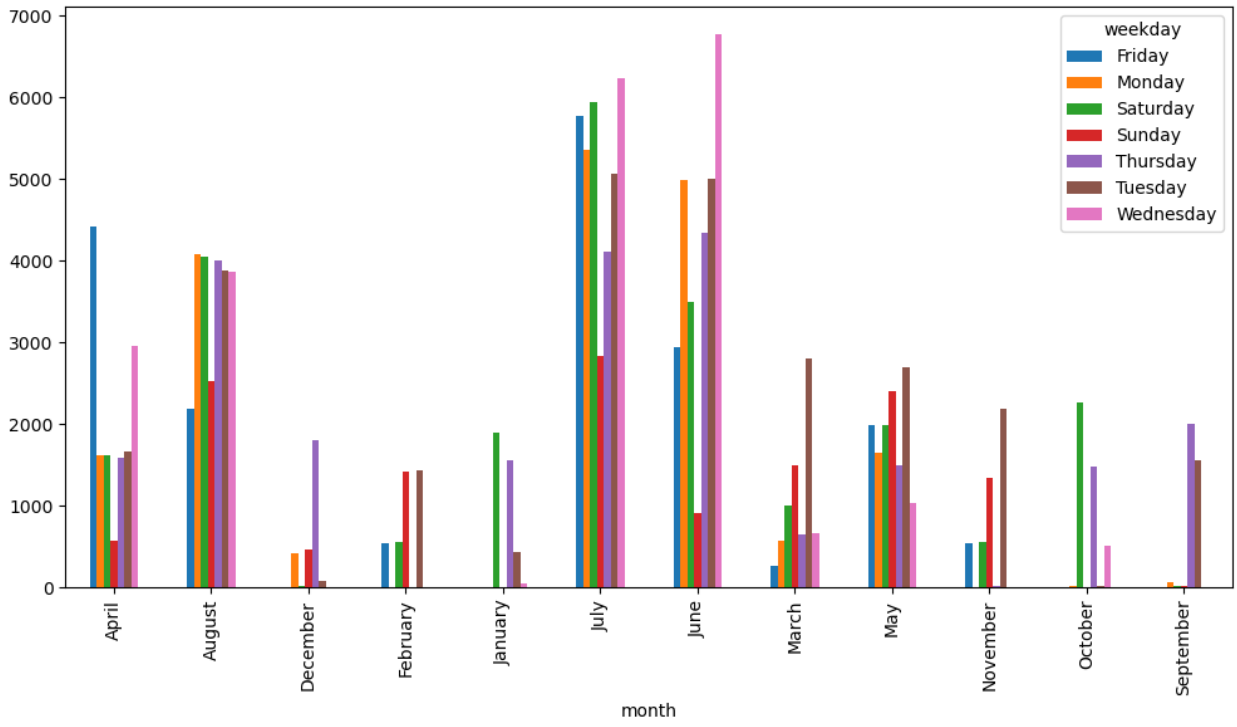
```
pivot = pd.crosstab(index=data['month'] , columns=data['weekday'])

pivot
```

weekday	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
month							
April	4415	1617	1622	564	1583	1667	2960
August	2178	4073	4040	2523	3997	3881	3859
December	0	416	16	457	1802	72	0
February	539	0	556	1412	0	1428	0
January	0	0	1895	0	1556	424	42

```
July          5772     5355       5938      2829       4104       5062
6230
June          2945     4987       3494       901       4342       5003
6772
March          262      573        999      1489        646       2804
666
May           1982     1645       1987      2397       1489       2695
1035
November       531        0        546      1334         13       2184
0
October          0       16       2260         0       1478         14
508
September        0       62         13        18       1998       1554
0
```

```
pivot.plot(kind='bar' , figsize=(12,6))
```

```
<Axes: xlabel='month'>
```



We cannot generalise infer about the Day of the week where their are the maximum leads generated , but we can generally infer from the bar graph that Tuesday and Saturdays are the days where we can assume that most of the leads are generated . So the team has to be very active on these 2 days particularly

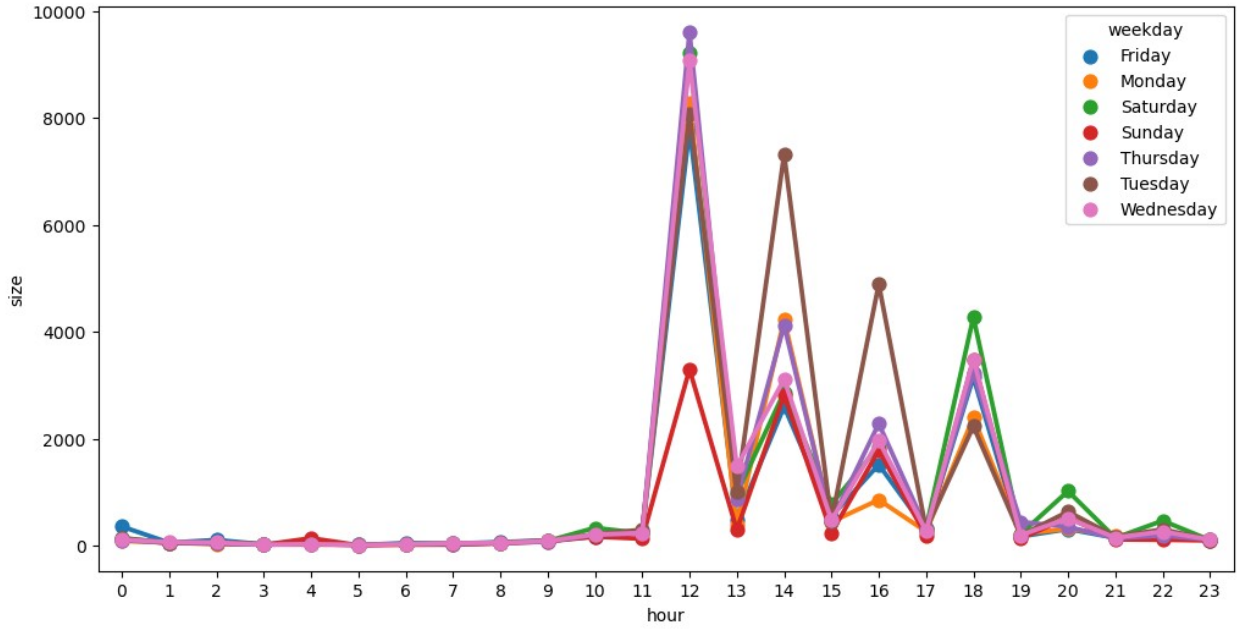
Hourly Rush on all days

```
summary = data.groupby(['weekday' , 'hour'] , as_index=False).size()

summary
```

```
       weekday  hour  size
0       Friday     0   359
1       Friday     1    56
2       Friday     2   108
3       Friday     3    25
4       Friday     4    68
..         ...   ...   ...
163  Wednesday    19   197
164  Wednesday    20   511
165  Wednesday    21   151
166  Wednesday    22   264
167  Wednesday    23   114

[168 rows x 3 columns]
```

```
import seaborn as sns
import matplotlib.pyplot as plt
plt.figure(figsize=(12,6))
sns.pointplot(x="hour" , y="size" , hue="weekday" , data=summary)
```

```
<Axes: xlabel='hour', ylabel='size'>
```



So we can infer from the graph that 12 P.M. followed by 2 P.M., 4 P.M. , 6 P.M. and 8 P.M. are the peak time where the least lead activity took place (eg where the last customer activity took place) . So these are the timings were the team has to be very very active during their work

```
data.head()
```

```
  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
6         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

    Lead Stage Lead Status Registration Device            Course  ...
\
0         Cold  Unverified             Unknown               BCA  ...

2    Untouched  Unverified             Unknown               BCA  ...

3         Cold  Unverified             Unknown               BCA  ...

4  No Response  Unverified             Unknown             B.Com  ...
```

```
6         Cold  Unverified             Unknown  B.Sc. Multimedia  ...


  Enrolment Status Organisation Campaign Number Channel Days to Last
Activity  \
0               No          ABC               89     API
0
2               No          ABC               89     API
0
3               No          ABC               89     API
0
4               No          ABC               89     API
0
6               No          ABC               89     API
0

    weekday day hour  minute month
0  Thursday  30   23      55  June
2  Thursday  30   23      50  June
3  Thursday  30   23      18  June
4  Thursday  30   22      57  June
6  Thursday  30   22      46  June

[5 rows x 24 columns]
```

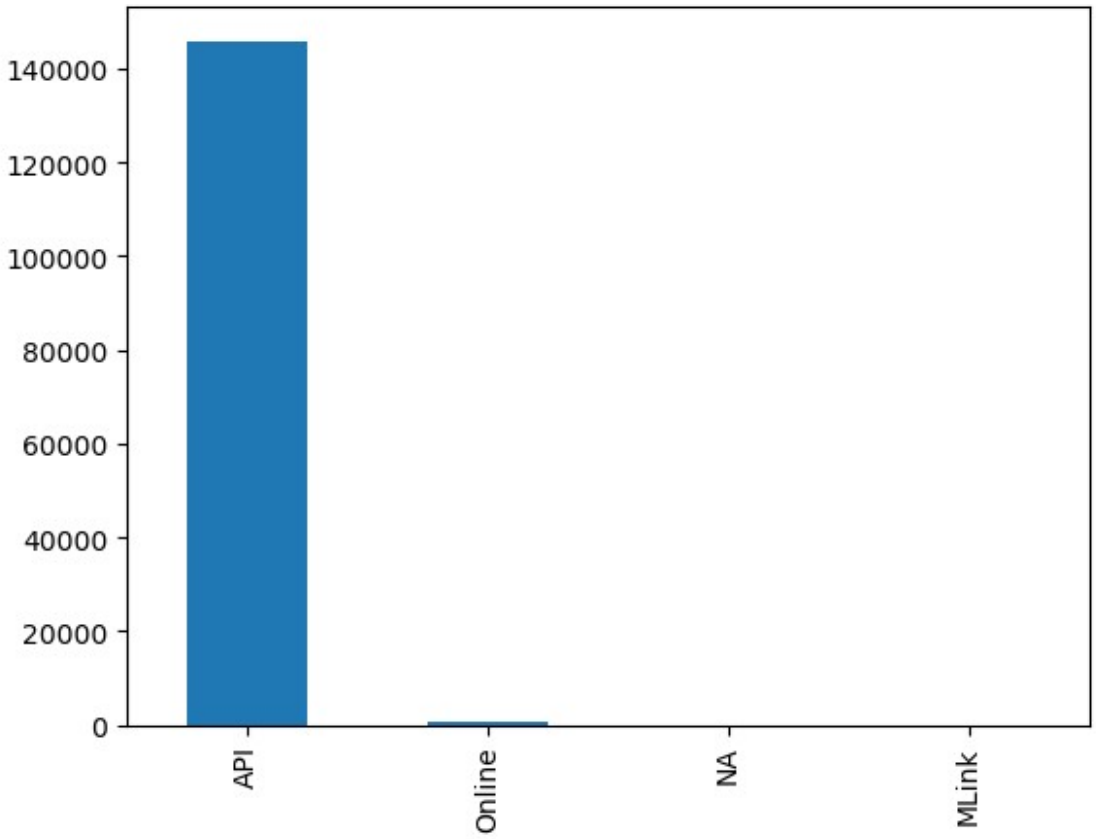
```
data.columns
```

```
Index(['Lead Origin', 'Country', 'State', 'City', 'Instance',
'Instance Date',
       'Lead Stage', 'Lead Status', 'Registration Device', 'Course',
'Campus',
       'Last Lead Activity Date', 'Form Initiated', 'Paid
Applications',
       'Enrolment Status', 'Organisation', 'Campaign Number',
'Channel',
       'Days to Last Activity', 'weekday', 'day', 'hour', 'minute',
'month'],
      dtype='object')
```

Which channel generates most number of Leads

```
data['Channel'].value_counts().plot(kind='bar')
```

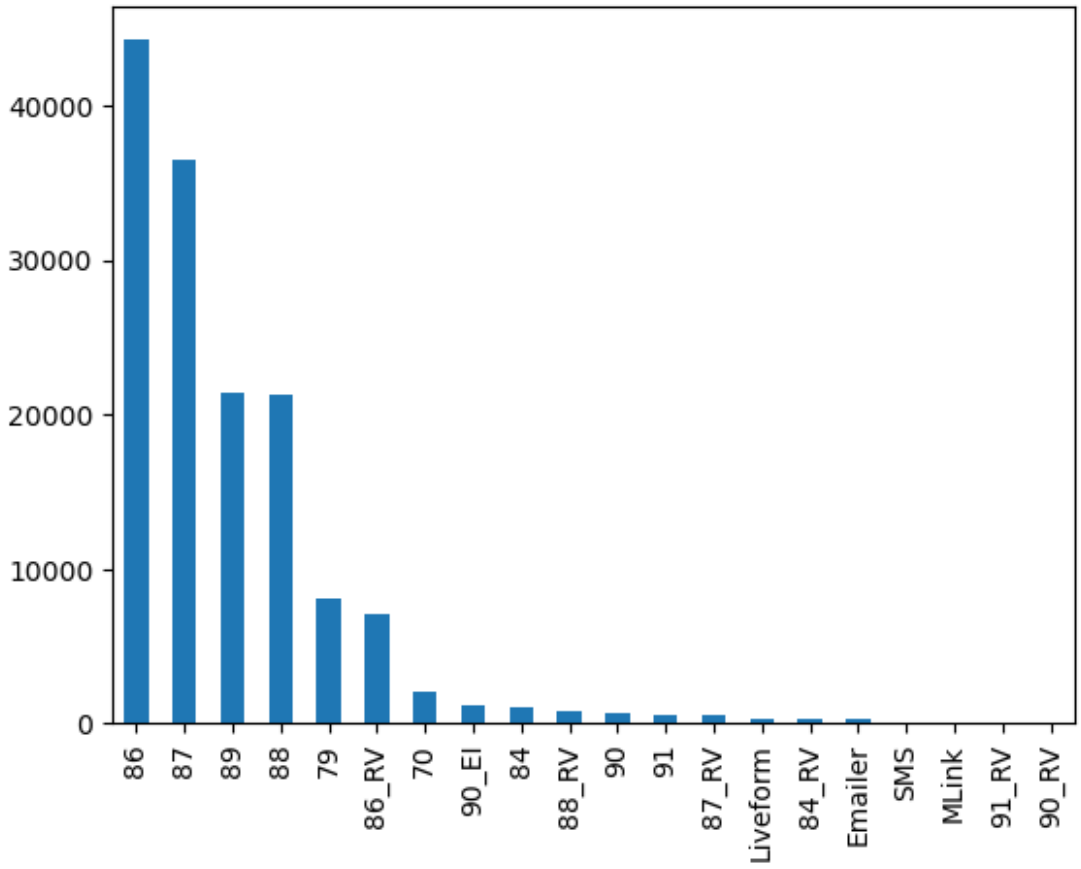
```
<Axes: >
```



Hence we can infer that most of the Leads that are generated are majorly through APIs following Online , Mlink channels . The difference is quite large . Hence we need to focus on more mediums of Lead generation espically in Online and MLink market place as they are quite unexplored areas and have untapped potential for more clients

```
data["Campaign Number"].value_counts().plot(kind='bar')
```

```
<Axes: >
```



86 Campaign Number generates most of the Leads . Followed by 87,89,88 and so on. Hence we need to keep emphasis on other Campaign Number also to actually hit the potential for the Lead generation in various campaigns

How long the most of the Leads lasted?

```
data['Days to Last Activity'].value_counts()
```

```
0      130484
1        1298
2         538
31        387
3         346
        ...
306         1
308         1
337         1
346         1
338         1
Name: Days to Last Activity, Length: 320, dtype: int64
```

Most of the Leads lasted less than 1 day(approx-> 1.3 Lakh) leads , this means that most of the Leads just came for a single point of information and had no intention of using any services of the company . Hence thos can be improved , either by increasing perks and features of the website or including more relevant services for the users to use

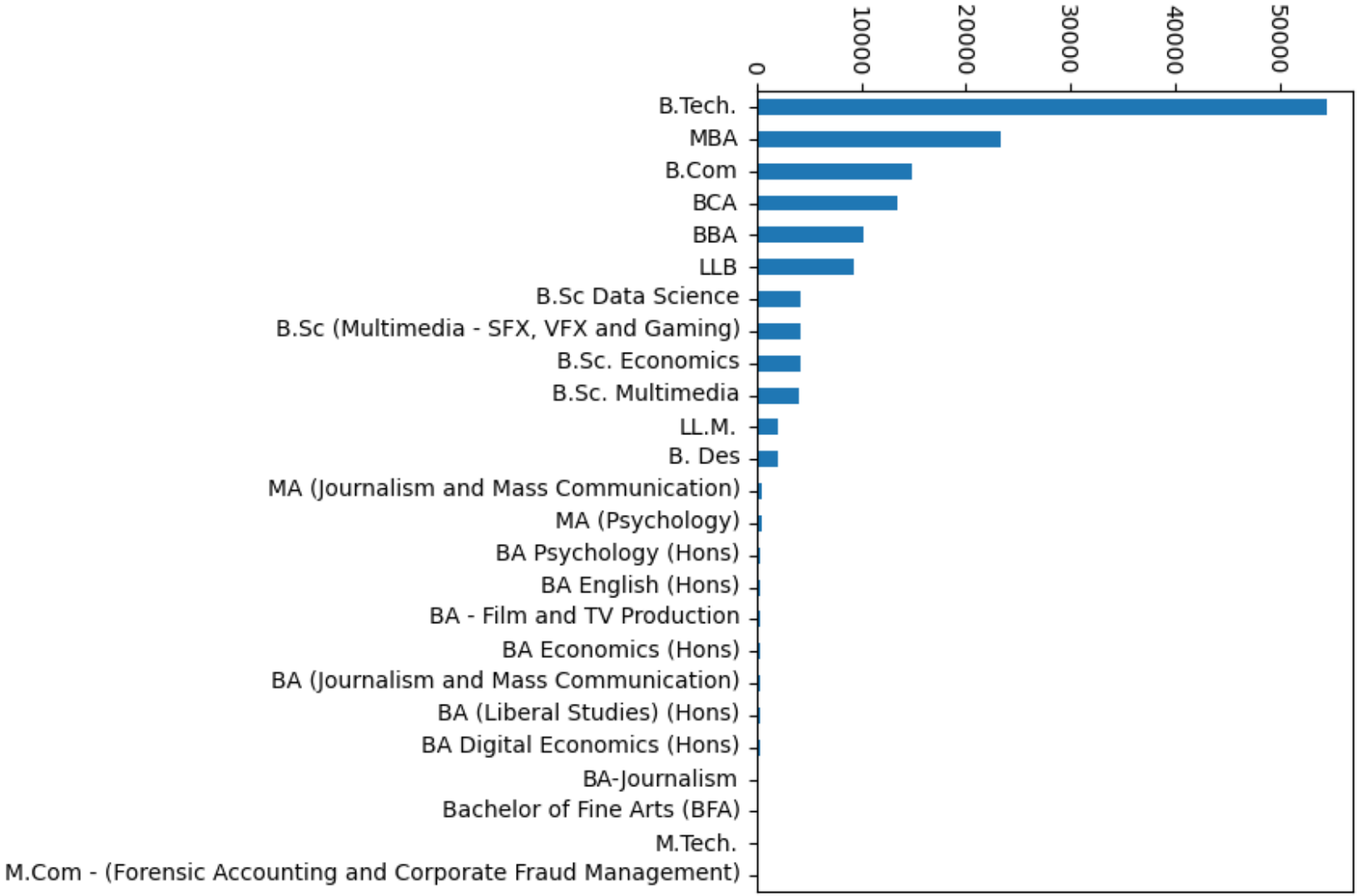
In which course do we have the maximum leads generated

```
data['Course'].value_counts()
```

```
B.Tech.                                              54300
MBA                                                  23106
B.Com                                                14630
BCA                                                  13237
BBA                                                  10126
LLB                                                   9181
B.Sc Data Science                                     4116
B.Sc (Multimedia - SFX, VFX and Gaming)               3999
B.Sc. Economics                                       3999
B.Sc. Multimedia                                      3937
LL.M.                                                 1916
B. Des                                                1915
MA (Journalism and Mass Communication)                 394
MA (Psychology)                                        391
BA Psychology (Hons)                                   202
BA English (Hons)                                      168
BA - Film and TV Production                            159
BA Economics (Hons)                                    155
BA (Journalism and Mass Communication)                 148
BA (Liberal Studies) (Hons)                            139
BA Digital Economics (Hons)                            136
BA-Journalism                                          119
Bachelor of Fine Arts (BFA)                             38
M.Tech.                                                 14
M.Com - (Forensic Accounting and Corporate Fraud Management)     1
Name: Course, dtype: int64
```

```
data['Course'].value_counts().plot(kind='bar')
```

```
<Axes: >
```


0
10000
20000
30000
40000
50000
B.Tech.
MBA
B.Com
BCA
BBA
LLB
B.Sc Data Science
B.Sc (Multimedia - SFX, VFX and Gaming)
B.Sc. Economics
B.Sc. Multimedia
LL.M.
B. Des
MA (Journalism and Mass Communication)
MA (Psychology)
BA Psychology (Hons)
BA English (Hons)
BA - Film and TV Production
BA Economics (Hons)
BA (Journalism and Mass Communication)
BA (Liberal Studies) (Hons)
BA Digital Economics (Hons)
BA-Journalism
Bachelor of Fine Arts (BFA)
M.Tech.
M.Com - (Forensic Accounting and Corporate Fraud Management)

We can infer that the Maximum leads generated are regarding the course Btech followed by MBA and Bcom. So we have to keep the systems of Btech , MBA and Bcom espically very fast and lucrative to convert those leads into clients. Side by side their is a strong upcoming interest shown by BA studies in various subjects. That is an untapped potential which we can increase or work for those markets also.

```
data.columns
```

```
Index(['Lead Origin', 'Country', 'State', 'City', 'Instance',
'Instance Date',
       'Lead Stage', 'Lead Status', 'Registration Device', 'Course',
'Campus',
       'Last Lead Activity Date', 'Form Initiated', 'Paid
Applications',
       'Enrolment Status', 'Organisation', 'Campaign Number',
'Channel',
       'Days to Last Activity', 'weekday', 'day', 'hour', 'minute',
'month'],
      dtype='object')
```

```
df = data[['Lead Stage', 'Lead Status']]

pivot_table = df.pivot_table(index='Lead Stage', columns='Lead
Status', aggfunc='size', fill_value=0)

pivot_table['Total'] = pivot_table.sum(axis=1)

pivot_table['Verified (%)'] = (pivot_table.get('Verified', 0) /
pivot_table['Total']) * 100
pivot_table['Unverified (%)'] = (pivot_table.get('Unverified', 0) /
pivot_table['Total']) * 100

print(pivot_table)
```

```
Lead Status            Unverified  Verified  Total  Verified (%)  \
Lead Stage
Application Submitted           6       483    489     98.773006
Cold                        17309     12838  30147     42.584668
Hot                            92       133    225     59.111111
Invalid                      1612       823   2435     33.798768
NOT ELIGIBLE                  260       296    556     53.237410
NOT INTERESTED               5607      3310   8917     37.120108
No Response                 16224      9712  25936     37.446021
Reject                         16        53     69     76.811594
Selected                        1         2      3     66.666667
Untouched                   74221         2  74223      0.002695
Warm                         1649      1877   3526     53.233125

Lead Status            Unverified (%)
Lead Stage
Application Submitted       1.226994
```

```
Cold                          57.415332
Hot                           40.888889
Invalid                       66.201232
NOT ELIGIBLE                  46.762590
NOT INTERESTED                62.879892
No Response                   62.553979
Reject                        23.188406
Selected                      33.333333
Untouched                     99.997305
Warm                          46.766875
```

```
pivot_table.style.background_gradient()
```

```
<pandas.io.formats.style.Styler at 0x126e2a4df10>
```

```
df2 = data[['Lead Stage', 'Enrolment Status']]

pivot_table2 = df2.pivot_table(index='Lead Stage', columns='Enrolment
Status', aggfunc='size', fill_value=0)

pivot_table2['Total'] = pivot_table2.sum(axis=1)

pivot_table2['Enrolled (%)'] = (pivot_table2.get('Yes', 0) /
pivot_table2['Total']) * 100
pivot_table2['Unenrolled (%)'] = (pivot_table2.get('No', 0) /
pivot_table2['Total']) * 100

print(pivot_table2)
```

```
Enrolment Status           No  Yes  Total  Enrolled (%)  Unenrolled (%)
Lead Stage
Application Submitted     299  190    489     38.854806       61.145194
Cold                    30147    0  30147      0.000000      100.000000
Hot                       225    0    225      0.000000      100.000000
Invalid                  2435    0   2435      0.000000      100.000000
NOT ELIGIBLE              556    0    556      0.000000      100.000000
NOT INTERESTED           8917    0   8917      0.000000      100.000000
No Response             25936    0  25936      0.000000      100.000000
Reject                     69    0     69      0.000000      100.000000
Selected                    3    0      3      0.000000      100.000000
Untouched               74223    0  74223      0.000000      100.000000
Warm                     3526    0   3526      0.000000      100.000000
```

```
pivot_table2.style.background_gradient()
```

```
<pandas.io.formats.style.Styler at 0x126e2a99690>
```

Their are number of Conclusions we can draw from these 2 pivot tables . They tell us about a large number of areas for improvement.

```
data.head()
```

```
  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
6         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

    Lead Stage Lead Status Registration Device            Course  ...
\
0         Cold  Unverified             Unknown               BCA  ...

2    Untouched  Unverified             Unknown               BCA  ...

3         Cold  Unverified             Unknown               BCA  ...

4  No Response  Unverified             Unknown             B.Com  ...

6         Cold  Unverified             Unknown  B.Sc. Multimedia  ...


  Enrolment Status Organisation Campaign Number Channel Days to Last
Activity  \
0               No          ABC              89     API
0
2               No          ABC              89     API
0
3               No          ABC              89     API
0
4               No          ABC              89     API
0
6               No          ABC              89     API
0

    weekday day hour  minute month
0  Thursday  30   23      55  June
2  Thursday  30   23      50  June
3  Thursday  30   23      18  June
4  Thursday  30   22      57  June
6  Thursday  30   22      46  June

[5 rows x 24 columns]
```

```
df2 = data[['Course', 'Enrolment Status']]
```

```
pivot_table3 = df2.pivot_table(index='Course', columns='Enrolment
Status', aggfunc='size', fill_value=0)

pivot_table3['Total'] = pivot_table3.sum(axis=1)

pivot_table3['Enrolled (%)'] = (pivot_table3.get('Yes', 0) /
pivot_table3['Total']) * 100
pivot_table3['Unenrolled (%)'] = (pivot_table3.get('No', 0) /
pivot_table3['Total']) * 100

pivot_table3.style.background_gradient()
```

```
<pandas.io.formats.style.Styler at 0x126e36297d0>
```

```
data.head()
```

```
  Lead Origin Country      State              City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
6         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

    Lead Stage Lead Status Registration Device             Course  ...
\
0         Cold  Unverified             Unknown                BCA  ...

2    Untouched  Unverified             Unknown                BCA  ...

3         Cold  Unverified             Unknown                BCA  ...

4  No Response  Unverified             Unknown              B.Com  ...

6         Cold  Unverified             Unknown  B.Sc. Multimedia  ...


  Enrolment Status Organisation Campaign Number Channel Days to Last
Activity  \
0               No          ABC              89     API
0
2               No          ABC              89     API
0
3               No          ABC              89     API
0
4               No          ABC              89     API
```

```
0
6                No           ABC                89     API
0
```

	weekday	day	hour	minute	month
0	Thursday	30	23	55	June
2	Thursday	30	23	50	June
3	Thursday	30	23	18	June
4	Thursday	30	22	57	June
6	Thursday	30	22	46	June

[5 rows x 24 columns]

State-wise Lead Generation

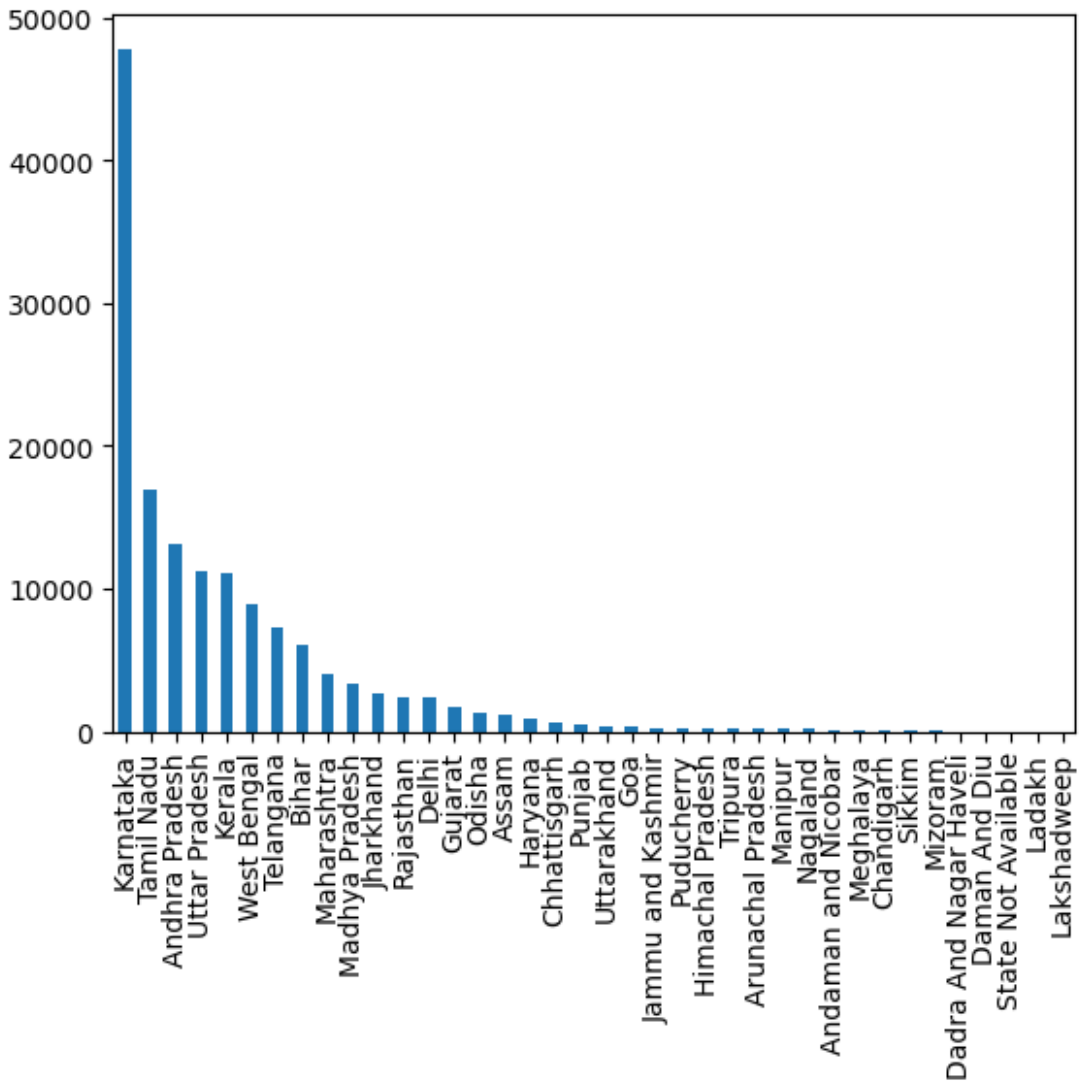
```
data["State"].value_counts()
```

```
Karnataka                 47816
Tamil Nadu                16991
Andhra Pradesh            13148
Uttar Pradesh             11285
Kerala                    11075
West Bengal                8913
Telangana                  7316
Bihar                      6054
Maharashtra                4095
Madhya Pradesh             3282
Jharkhand                  2732
Rajasthan                  2451
Delhi                      2338
Gujarat                    1660
Odisha                     1368
Assam                      1197
Haryana                     886
Chhattisgarh                641
Punjab                      443
Uttarakhand                 407
Goa                         404
Jammu and Kashmir           292
Puducherry                  259
Himachal Pradesh            202
Tripura                     201
Arunachal Pradesh           195
Manipur                     185
Nagaland                    161
Andaman and Nicobar         139
Meghalaya                   133
Chandigarh                   95
Sikkim                       72
```

```
Mizoram                          56
Dadra And Nagar Haveli           17
Daman And Diu                     8
State Not Available               6
Ladakh                            2
Lakshadweep                       1
Name: State, dtype: int64
```

```
data["State"].value_counts().plot(kind='bar')
```

```
<Axes: >
```



Maximum Leads were generated from Karnataka followed by Tamil Nade, Andra Pradesh and Uttar Pradesh

```
df3 = data[['State', 'Enrolment Status']]

pivot_table4 = df3.pivot_table(index='State', columns='Enrolment
Status', aggfunc='size', fill_value=0)

pivot_table4['Total'] = pivot_table4.sum(axis=1)

pivot_table4['Enrolled (%)'] = (pivot_table4.get('Yes', 0) /
pivot_table4['Total']) * 100
pivot_table4['Unenrolled (%)'] = (pivot_table4.get('No', 0) /
pivot_table4['Total']) * 100

pivot_table4.sort_values(by=['Enrolled
(%)']).style.background_gradient()
```

```
<pandas.io.formats.style.Styler at 0x126e365b790>
```

```
final.head()
```

```
  Lead Origin Country      State             City Instance Instance
Date  \
0         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
1         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
2         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
3         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30
4         API   India  Karnataka  Bengaluru Rural  Primary    2022-06-
30

       Lead Stage Lead Status Registration Device            Course  \
0            Cold  Unverified             Unknown               BCA
1  NOT INTERESTED  Unverified             Unknown  B.Sc. Multimedia
2       Untouched  Unverified             Unknown               BCA
3            Cold  Unverified             Unknown               BCA
4     No Response  Unverified             Unknown              B.Com

                     Campus Last Lead Activity Date Form Initiated  \
0  University Not Available     2022-06-30 23:55:00  Not Initiated
1  University Not Available     2022-01-07 10:07:00  Not Initiated
2  University Not Available     2022-06-30 23:50:00  Not Initiated
3  University Not Available     2022-06-30 23:18:00  Not Initiated
4  University Not Available     2022-06-30 22:57:00  Not Initiated

  Paid Applications Enrolment Status Organisation Campaign Number
Channel  \
```

```
0          Not Paid                No          ABC                89
API
1          Not Paid                No          ABC                89
API
2          Not Paid                No          ABC                89
API
3          Not Paid                No          ABC                89
API
4          Not Paid                No          ABC                89
API
```

```
   Days to Last Activity
0                      0
1                   -174
2                      0
3                      0
4                      0
```

```
final["City"]
```

```
0            Bengaluru Rural
1            Bengaluru Rural
2            Bengaluru Rural
3            Bengaluru Rural
4            Bengaluru Rural
                 ...
149995    South 24 Parganas
149996              Kalyani
149997              Kolkata
149998              Kolkata
149999              Kolkata
Name: City, Length: 150000, dtype: object
```

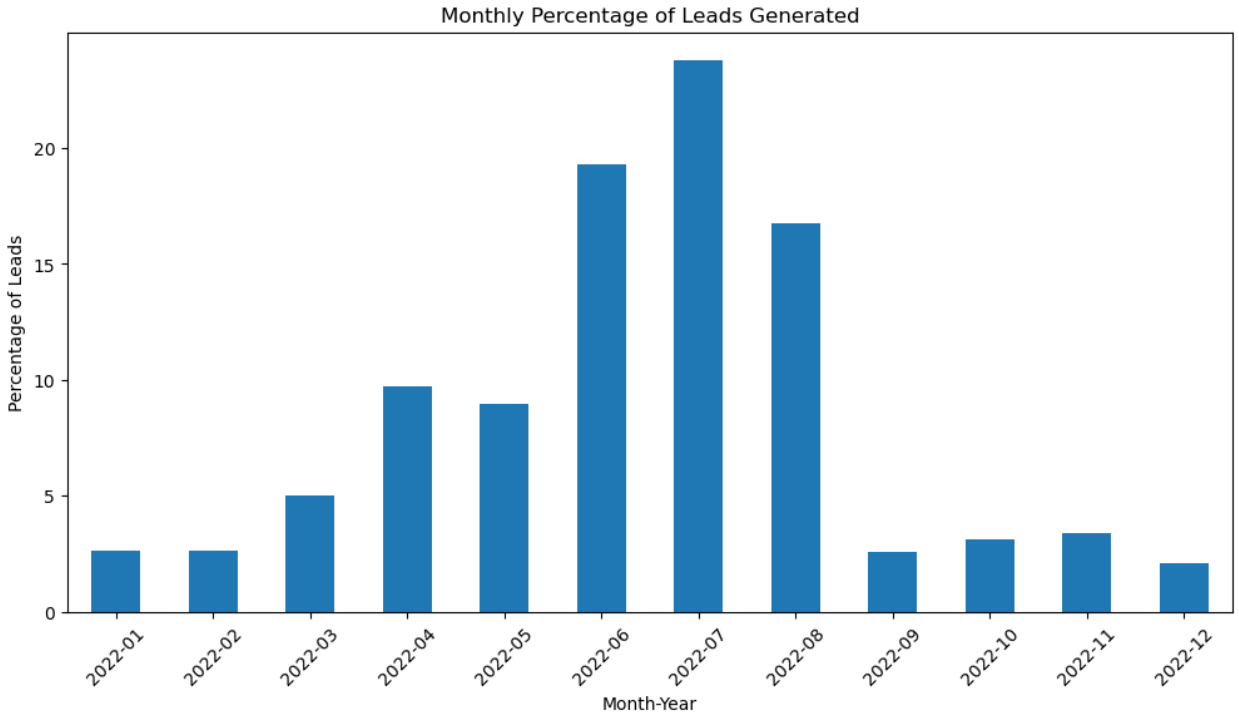
```
min_instance_date = final['Instance Date'].min()
max_instance_date = final['Instance Date'].max()

min_last_activity_date = final['Last Lead Activity Date'].min()
max_last_activity_date = final['Last Lead Activity Date'].max()

# Display the results
print(f"Instance Date Range: {min_instance_date} to
{max_instance_date}")
print(f"Last Lead Activity Date Range: {min_last_activity_date} to
{max_last_activity_date}")
```

```
Instance Date Range: 2022-01-04 00:00:00 to 2022-12-08 00:00:00
Last Lead Activity Date Range: 1970-01-01 05:30:00 to 2023-04-01
13:17:00
```

```
final['YearMonth'] = final['Instance Date'].dt.to_period('M')
```

```
monthly_leads = final.groupby('YearMonth').size()

total_leads = monthly_leads.sum()
monthly_percentage = (monthly_leads / total_leads) * 100

monthly_percentage_df =
monthly_percentage.reset_index(name='Percentage')

print(monthly_percentage_df)

import matplotlib.pyplot as plt

monthly_percentage.plot(kind='bar', figsize=(12, 6))
plt.title('Monthly Percentage of Leads Generated')
plt.xlabel('Month-Year')
plt.ylabel('Percentage of Leads')
plt.xticks(rotation=45)
plt.show()
```

	YearMonth	Percentage
0	2022-01	2.611333
1	2022-02	2.626667
2	2022-03	5.010667
3	2022-04	9.746000
4	2022-05	8.966000
5	2022-06	19.303333
6	2022-07	23.803333
7	2022-08	16.737333
8	2022-09	2.574000
9	2022-10	3.101333
10	2022-11	3.420000
11	2022-12	2.100000




```
(final["State"].value_counts()/150000)* 100
```

```
Karnataka                32.780667
Tamil Nadu               11.641333
Andhra Pradesh            8.949333
Uttar Pradesh             7.640000
Kerala                    7.517333
West Bengal               6.114000
Telangana                 4.980667
Bihar                     4.088667
Maharashtra               2.785333
Madhya Pradesh            2.225333
Jharkhand                 1.869333
Rajasthan                 1.660000
Delhi                     1.579333
Gujarat                   1.140667
Odisha                    0.934000
Assam                     0.816667
Haryana                   0.601333
Chhattisgarh              0.438667
Punjab                    0.304000
Uttarakhand               0.278000
Goa                       0.274000
Jammu and Kashmir         0.198667
Puducherry                0.176667
Tripura                   0.138667
Himachal Pradesh          0.136000
```

```
Arunachal Pradesh          0.132000
Manipur                    0.126000
Nagaland                   0.109333
Andaman and Nicobar        0.094667
Meghalaya                  0.092000
Chandigarh                 0.065333
Sikkim                     0.050000
Mizoram                    0.037333
Dadra And Nagar Haveli     0.012000
Daman And Diu              0.005333
State Not Available        0.005333
Ladakh                     0.001333
Lakshadweep                0.000667
Name: State, dtype: float64
```

```
(final["City"].value_counts()/150000)* 100
```

```
Bengaluru Rural    25.925333
Chennai             5.308667
Hyderabad           4.108000
Kolkata             3.838667
New Delhi           1.575333
                     ...
Tenga Market        0.000667
Mudargi             0.000667
Kanhangad           0.000667
Yadgir              0.000667
Dapoli              0.000667
Name: City, Length: 1167, dtype: float64
```

```
final["City"].unique()
```

```
array(['Bengaluru Rural', 'Bengaluru', 'Kanyakumari', ..., 'Manbazar',
       'Kadegaon', 'Dapoli'], dtype=object)
```

```
(data["weekday"].value_counts()/150000)*100
```

```
Tuesday      17.858667
Saturday     15.577333
Thursday     15.338667
Wednesday    14.714667
Monday       12.496000
Friday       12.416000
Sunday        9.282667
Name: weekday, dtype: float64
```

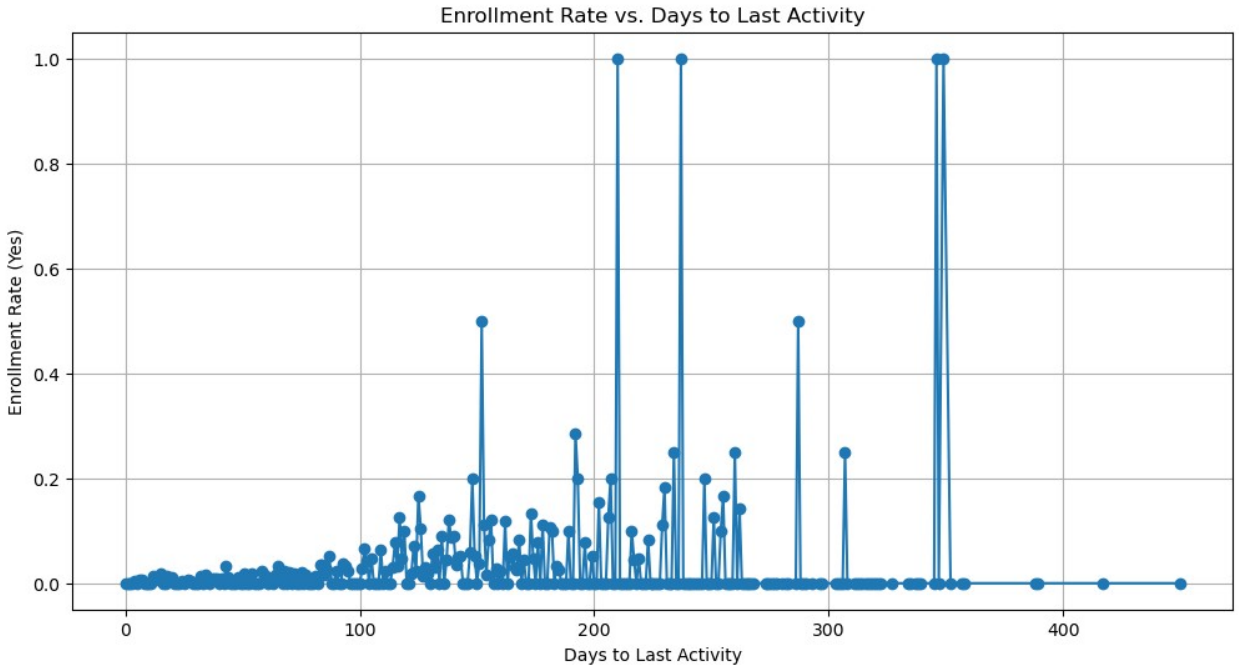
```
(data["hour"].value_counts()/150000)*100
```

```
12    36.905333
14    18.052000
18    14.839333
```

```
16    10.104000
13     3.668000
20     2.452667
15     2.272667
17     1.326000
22     1.154667
11     1.081333
19     1.050000
10     0.988000
0      0.711333
21     0.708667
23     0.498000
9      0.405333
2      0.268667
1      0.257333
8      0.248667
4      0.234000
7      0.155333
6      0.134667
3      0.100000
5      0.068000
Name: hour, dtype: float64
```

```
df_filtered = data[['Days to Last Activity', 'Enrolment Status']]

df_filtered['Enrolment Status'] = df_filtered['Enrolment
Status'].map({'Yes': 1, 'No': 0})

grouped_df = df_filtered.groupby('Days to Last Activity')['Enrolment
Status'].mean().reset_index()

plt.figure(figsize=(12, 6))
plt.plot(grouped_df['Days to Last Activity'], grouped_df['Enrolment
Status'], marker='o', linestyle='-')
plt.title('Enrollment Rate vs. Days to Last Activity')
plt.xlabel('Days to Last Activity')
plt.ylabel('Enrollment Rate (Yes)')
plt.grid(True)
plt.show()
```

```
C:\Users\Yashita\AppData\Local\Temp\ipykernel_14064\556862389.py:3:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  df_filtered['Enrolment Status'] = df_filtered['Enrolment
Status'].map({'Yes': 1, 'No': 0})
```




```
df_filtered = data[['Days to Last Activity', 'Enrolment Status']]

pivot_table = df_filtered.pivot_table(index='Days to Last Activity',
columns='Enrolment Status', aggfunc='size', fill_value=0)

print(pivot_table.columns)

# Add a total count column
pivot_table['Total'] = pivot_table.sum(axis=1)

yes_column = 'Yes'
no_column = 'No'

pivot_table['Enrolled (%)'] = (pivot_table[yes_column] /
pivot_table['Total']) * 100
pivot_table['Unenrolled (%)'] = (pivot_table[no_column] /
pivot_table['Total']) * 100

print(pivot_table)
```

```
Index(['No', 'Yes'], dtype='object', name='Enrolment Status')
Enrolment Status           No  Yes   Total  Enrolled (%)  Unenrolled
(%)
Days to Last Activity

0                      130484    0  130484      0.000000
100.000000
1                        1298    0    1298      0.000000
100.000000
2                         538    0     538      0.000000
```

```
100.000000
3                          346    0     346      0.000000
100.000000
4                          251    1     252      0.396825
99.603175
...                        ...  ...     ...           ...               .
..
358                          2    0       2      0.000000
100.000000
388                          1    0       1      0.000000
100.000000
389                          2    0       2      0.000000
100.000000
417                          1    0       1      0.000000
100.000000
450                          1    0       1      0.000000
100.000000

[320 rows x 5 columns]
```