

Robot Operating System: How to pitch a Robot's Brain?

A DEEP DIVE ON EFFECTIVE MARKET CAMPAIGN
STRATEGIES FOR HIGHER ROI ON THE ROS

TABLE OF CONTENTS

01 Case study - A brief summary

02 Data description

03 Data pre-processing strategies

04 Exploratory Data Analysis

05 Statistical testing

06 Insights & Recommendations

07 Data challenges

08 Conclusion

“Key considerations when choosing a robot’s operating system”

“Securing ROS robotics platforms”

“Docker & ROS: When all you have is a hammer, everything looks like a nail”

“ROS Support”

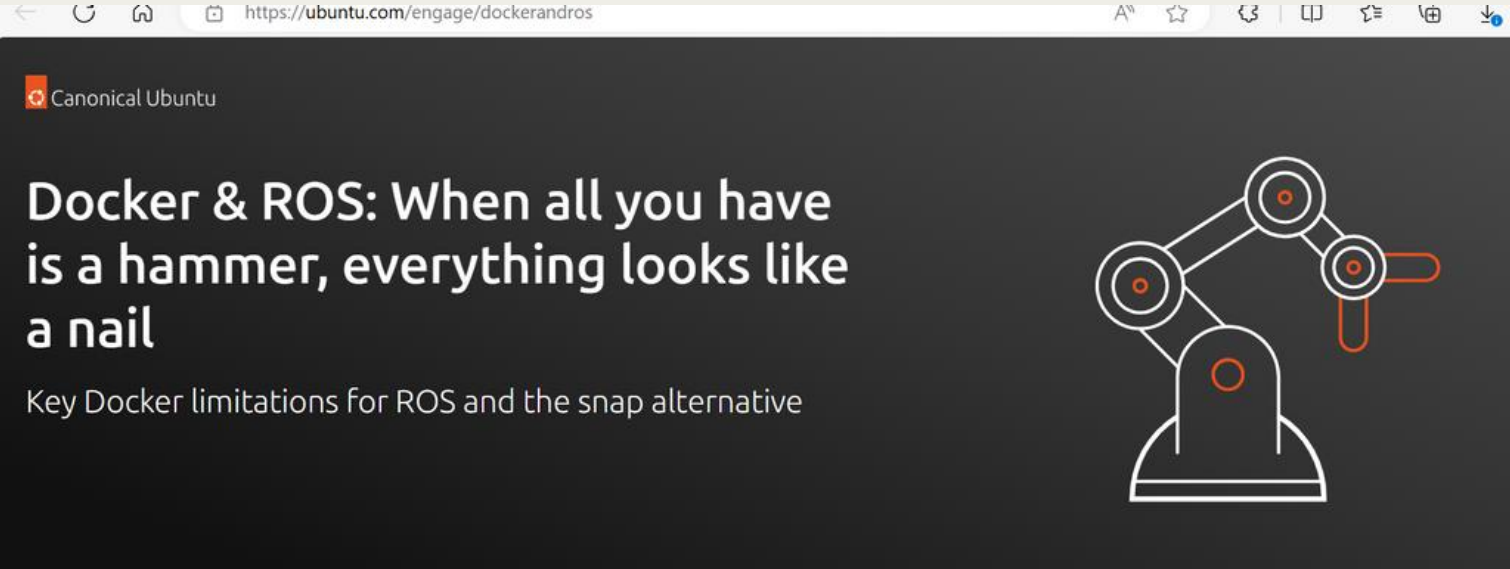
MEET THE CAMPAIGNS



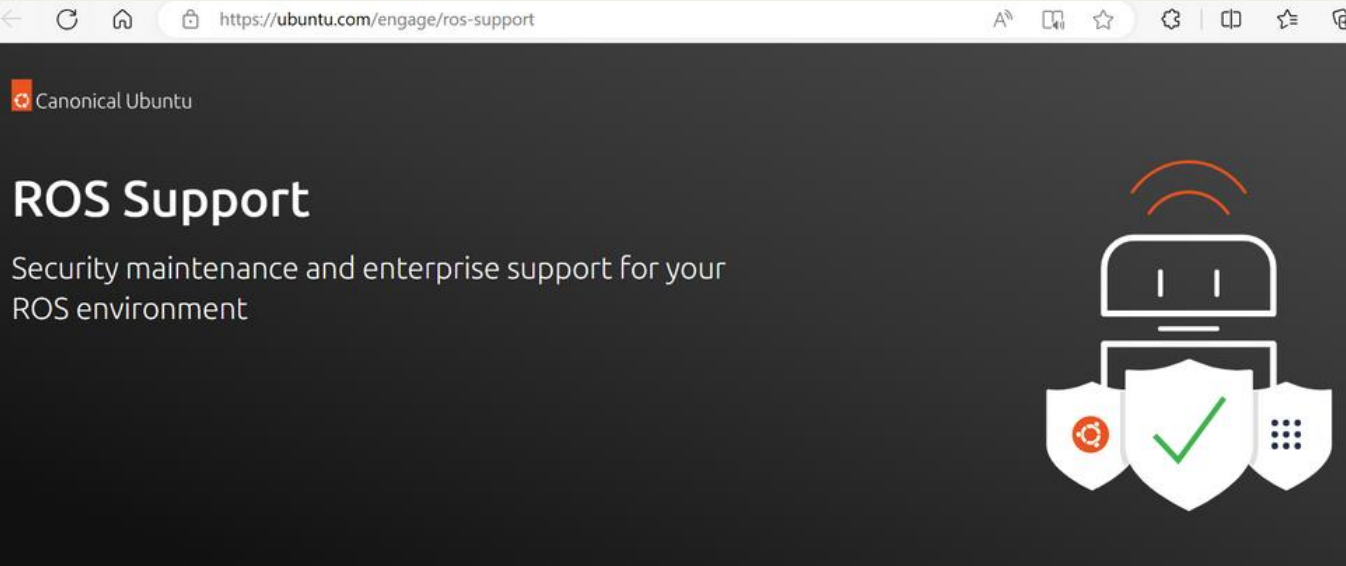
C1 - FY19_IOT_Robotics_Whitepaper_OSconsiderations



C2 - CY20_IOT_Robotics_Whitepaper_ROSonrobotics



C3 - CY21_IOT_Robotics_T1_WP_ROSDocker



C4 - CY21_IOT_Robotics_Whitepaper_ROSESM

BACKGROUND

- Explore the effectiveness and outreach of 4 marketing campaigns with respect to a Robot Operating System whitepaper.
- Aims to establish standard metrics of evaluation for gauging campaign effectiveness.
- Uncover primary factors that influence market growth in terms of views, leads and conversions for each campaign strategy.
- Suggest strategies to improve gray areas in data collection to enhance future campaign planning.

A GLANCE AT THE DATA

- 3 datasets for 4 marketing campaigns in form of Campaigns, Pages and Leads.
- Each row of Pages dataset corresponds to page related information (e.g. – when a *user viewed a campaign page*, no of times a page link is visited, etc.)
- Each row of the Leads dataset corresponds to leads’ specific details (e.g. – when a *user joined a campaign*, lead status, campaign name, etc.)
- 9.2% data in foreign languages (lead_job_title, industry, source, country, etc.)
- 50,410 total page views across 354 distinct page viewing dates for all campaigns.
- 2540 total leads spanning 263 unique campaign joining dates from 117 countries for all campaigns.
- Missing data ranges from 1% to 80% in some columns, after cleaning and merging – 2507 rows and 14 columns (1.3% loss, 2540 rows earlier).

```
marketing_campaigns_merged_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2507 entries, 0 to 2506
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  -
0   campaign_name          2507 non-null   object
1   total_page_views       2507 non-null   float64
2   page_avg_time          2507 non-null   timedelta64[ns]
3   page_bounce_rate       2507 non-null   float64
4   view_date_range        2507 non-null   object
5   page_avg_views         2507 non-null   float64
6   lead_hashed_id         2507 non-null   object
7   campaign_joined_date   2507 non-null   datetime64[ns]
8   lead_job_title         2507 non-null   object
9   lead_country           2507 non-null   object
10  lead_source            2507 non-null   object
11  lead_status            2507 non-null   object
12  weekday                2507 non-null   object
13  month                  2507 non-null   object
dtypes: datetime64[ns](1), float64(3), object(9), timedelta64[ns](1)
memory usage: 274.3+ KB
```


DATA PREPROCESSING STRATEGIES

```
#checking the value of each type of job titles
marketing_campaigns_merged_df['lead_job_title'].value
```

	count
lead_job_title	
student	84
Student	65
CEO	48
1	36
Engineer	36

Data cleaning & prep

- Merging 3 datasets on basis of Campaign_name field ensuring no duplication.
- Removed commas and unnecessary characters from campaign_names.
- Corrected inconsistent values (‘Doctor’ to ‘Dr.’, ‘STUDENT’ to ‘Student’, etc).
- 9.17% of foreign language texts.

```
#check percentage of missing values in the dataset
missing_values_percentage = (marketing_campaigns_merged_df.isnull().sum() / len(marketing_campaigns_merged_df)) * 100

#Give a label to the '0' in the 'missing_values_percentage'
missing_values_percentage.name = 'Missing Values Percentage'
missing_values_percentage
```

	Missing Values Percentage
campaign_name	0.000000
page_url	0.000000
page_view_date	0.000000
page_views	0.000000
page_avg_time	0.000000
page_bounce_rate	0.000000
lead_hashed_id	0.000000
campaign_joined_date	0.000000
lead_job_title	1.418897
lead_industry	80.050393
lead_country	14.896456
lead_source	36.287025
lead_status	0.000000

dtype: float64

Missing values

- Missing value columns all from LEADS dataset - lead_job_title, country, source, industry.
- Dropped lead_industry column with 85% missing data.
- Removed 1.3% of rows with ‘NaN’ for lead_job_title column.
- Imputed with ‘Unknown’ value for country and source columns with 15-36% missing fields.

```
[ ] #Convert page_views and page_bounce_rate into float data type
pages['page_views'] = pages['page_views'].astype(float)
pages['page_bounce_rate'] = pages['page_bounce_rate'].astype(float)
pages.info()
```

```
<<class 'pandas.core.frame.DataFrame'>
RangeIndex: 791 entries, 0 to 790
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   page_url    790 non-null    object
1   date        790 non-null    datetime64[ns]
2   page_views  791 non-null    float64
3   page_avg_time  791 non-null    object
4   page_bounce_rate  791 non-null    float64
dtypes: datetime64[ns](1), float64(2), object(2)
memory usage: 31.0+ KB
```

We will also convert the page_avg_time into a timedelta column so that we can carry out various time calculations on this column.

```
[ ] #convert page_avg_time into timedelta data type
pages['page_avg_time'] = pd.to_timedelta(pages['page_avg_time'])
pages.info()
```

```
<<class 'pandas.core.frame.DataFrame'>
RangeIndex: 791 entries, 0 to 790
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   page_url    790 non-null    object
1   date        790 non-null    datetime64[ns]
2   page_views  791 non-null    float64
3   page_avg_time  791 non-null    timedelta64[ns]
4   page_bounce_rate  791 non-null    float64
dtypes: datetime64[ns](1), float64(2), object(1), timedelta64[ns](1)
memory usage: 31.0+ KB
```

Format datatypes

- Modified the date columns to have DateTime format.
- Converted columns with Time information into TimeDelta datatype (eg- page_avg_time).
- Converted specific Object cols to appropriate types.

```
marketing_campaigns_merged_df.info()
```

```
<<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2507 entries, 0 to 2506
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  -
0   campaign_name    2507 non-null    object
1   total_page_views  2507 non-null    float64
2   page_avg_time    2507 non-null    timedelta64[ns]
3   page_bounce_rate  2507 non-null    float64
4   view_date_range  2507 non-null    object
5   page_avg_views    2507 non-null    float64
6   lead_hashed_id    2507 non-null    object
7   campaign_joined_date  2507 non-null    datetime64[ns]
8   lead_job_title    2507 non-null    object
9   lead_country      2507 non-null    object
10  lead_source       2507 non-null    object
11  lead_status       2507 non-null    object
12  weekday           2507 non-null    object
13  month             2507 non-null    object
dtypes: datetime64[ns](1), float64(3), object(9), timedelta64[ns](1)
memory usage: 274.3+ KB
```

Feature Engineering

- Added ‘weekday’ and ‘months’, extracted from Date columns.
- Dropped page_url column from merged dataset.
- Added numerical columns for specific categorical columns for statistical analysis part.

EXPLORATORY DATA ANALYSIS

Important Questions We Want To Answer:-

- How long do people spend on a page on an average before bouncing off?
- Which industries are producing the highest leads?
- How frequently do we observe days with high numbers of conversions compared to days with low numbers of conversions?
- Do more page views lead to more conversions?
- Which marketing campaign is more effective in terms of conversions and effectiveness?
- How many possible conversions did each lead source category result in?

EXPLORATORY DATA ANALYSIS-FINDINGS

CY20_IOT_Robotics_Whitepaper_ROSonrobotics	Nurture	91
	Disqualified	60
	New	42
	Unsubscribed	26
	MQL	8
	Contact	2
CY21_IOT_Robotics_T1_WP_ROSDocker	SQL	1
	Opportunity	1
	Disqualified	873
	Nurture	456
	Lead	105
	Unsubscribed	1

as an average of total_page_views for each campaign in the aggregated dataframe					
] = aggregated_page_data['total_page_views'] / 4					
ign_name	total_page_views	page_avg_time	page_bounce_rate		
nrobotics	1610.0	0 days 00:01:52.437288135	59.327898	2021-01-01 00:00:00	
SDocker	37122.0	0 days 00:01:13.875000	78.965500	2021-11-22 00:00:00	
_ROSESM	9630.0	0 days 00:01:27.571428571	49.142925	2021-07-05 00:00:00	
derations	2048.0	0 days 00:01:08.165584415	51.219058	2021-01-01 00:00:00	
age_data	View recommended plots		New interactive sheet		

bounce rate for each campaign	
ign_page_data.groupby('campaign_name')['page_bounce_rate'].mean().sort_values(asc	
	page_bounce_rate
campaign_name	
CY21_IOT_Robotics_T1_WP_ROSDocker	78.965500
0_IOT_Robotics_Whitepaper_ROSonrobotics	59.327898
_IOT_Robotics_Whitepaper_OSconsiderations	51.219058
Y21_IOT_Robotics_Whitepaper_ROSESM	49.142925
float64	

#1. How many users or leads joined each campaign?	
leads.groupby('campaign_name')['lead_hashed_id'].count().sort_values(ascending=False	
	lead_hashed_id
campaign_name	
CY21_IOT_Robotics_T1_WP_ROSDocker	1539
CY21_IOT_Robotics_Whitepaper_ROSESM	495
FY19_IOT_Robotics_Whitepaper_OSconsiderations	242
CY20_IOT_Robotics_Whitepaper_ROSonrobotics	231
dtype: int64	

Top lead statuses

Nurture, Disqualified

- The top statuses were ‘Nurture’ and ‘Disqualified’ for all the 4 campaigns, with **CY20_IOT_Robotics_Whitepaper_ROSonrobotics** and **CY21_IOT_Robotics_T1_WP_ROSDocker** being the only campaigns with ‘Sales Qualified Leads’ (SQL) count in the tally.

Avg page views/time

Views- 402-9280, Time- 1.52 sec

- Campaign 3 or CY21_IOT_Robotics_T1_WP_ROSDocker received highest average views of 9280 views, and c2 performed the worst with 402 views.
- CY20_IOT_Robotics_Whitepaper_ROSonrobotics had the highest average page view time of 1.52 mins, and c1 being the worst with 1.08 mins.

Average bounce rates

Range - 49 - 79%

- On an average, **59.66%** of visitors left the page without interacting further.
- CY21_IOT_Robotics_T1_WP_ROSDocker has the highest bounce rate of 79%, and CY21_IOT_Robotics_Whitepaper_ROSESM has the lowest with 49%.

Highest Leads

Campaign wise analysis

- Highest leads count (1539) for c3 - **CY21_IOT_Robotics_T1_WP_ROSDocker**, and lowest for CY20_IOT_Robotics_Whitepaper_ROSonrobotics with 231 leads.

SOME NUMBERS...

Nurture, Disqualified
top lead statuses

37122 views
#1C – ROSDock

9 unique
lead sources

41 distinct
leads’ industries

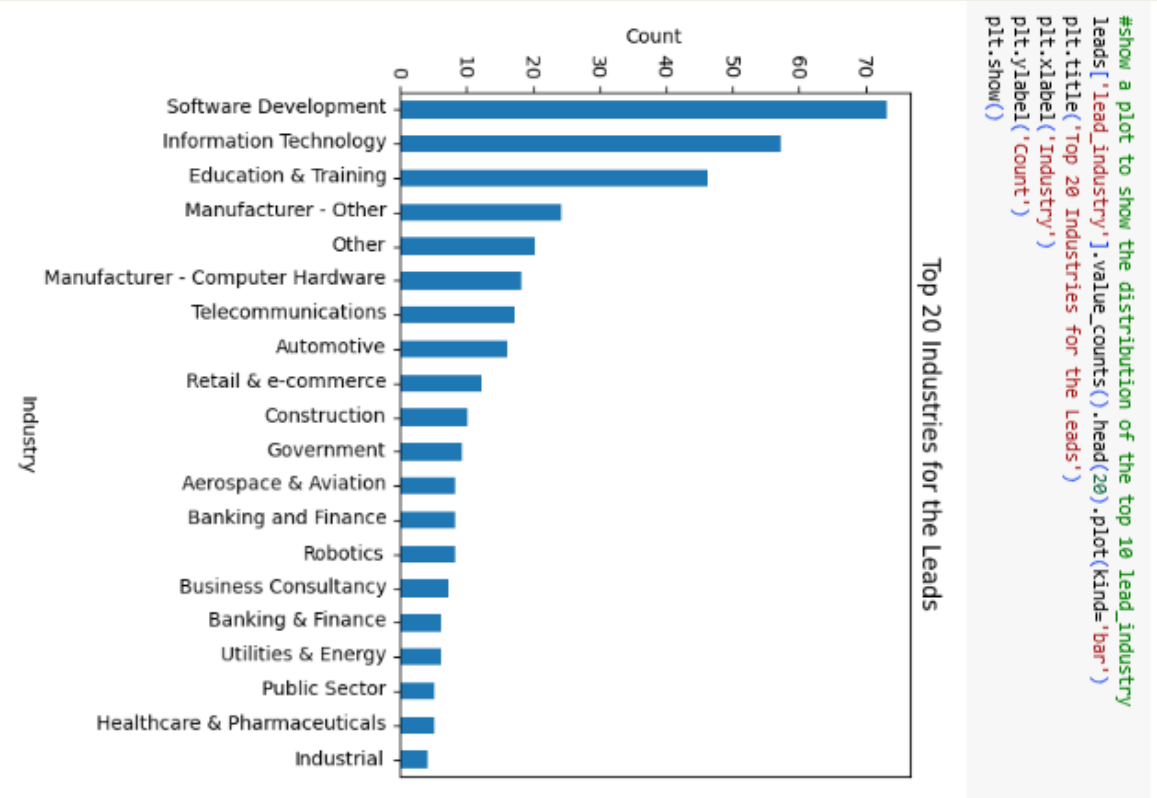
Google Paid
Top lead source

59.66%
overall page bounce rate

1539 highest leads
#1C – ROSDock

Software Development
Industry with highest leads

count	
lead_source	
Google paid	872
Unknown	627
Takeover	544
Email	205
Twitter organic	91
LinkedIn organic	86
Facebook organic	72
LinkedIn paid	5
Facebook paid	4



STATISTICAL TESTING - USE CASES

Hypothesis Testing

Chi-Squared test to determine if there is a **statistically significant relationship** between the performances of the campaigns.

- H0 (null hypothesis) - There is no difference between the performance and effectiveness of the 4 campaigns.
- H1 (alternate hypothesis) - The 4 campaigns vary in performance in terms of effectiveness.

Cohort Analysis

Segmentation analysis by grouping leads (cohorts) by 'campaign_joined_date' field to assess campaign performance as per seasonality trends.

- Time-based cohorts: Group customers based on when they first joined the campaign (e.g., campaign_joined_date).
- Behavior-based cohorts: Group leads based on their initial behavior (e.g., lead source, lead_job_title).
- Analyzed metrics over different time intervals (e.g., weekly, monthly in this case).

A/B testing or Causal Analysis

Experiment with different campaign elements to optimize results. Limited to real experimentation.

- Divide campaigns into 4 groups and run the A/B test for a long enough period to collect enough data.
- For multiple comparisons, performed the Bonferroni correction to adjust the significance level (p-value).
- Mann-Whitney U Test and chi squared tests for non-parametric values.

CAMPAIGN PERFORMANCE ANALYSIS

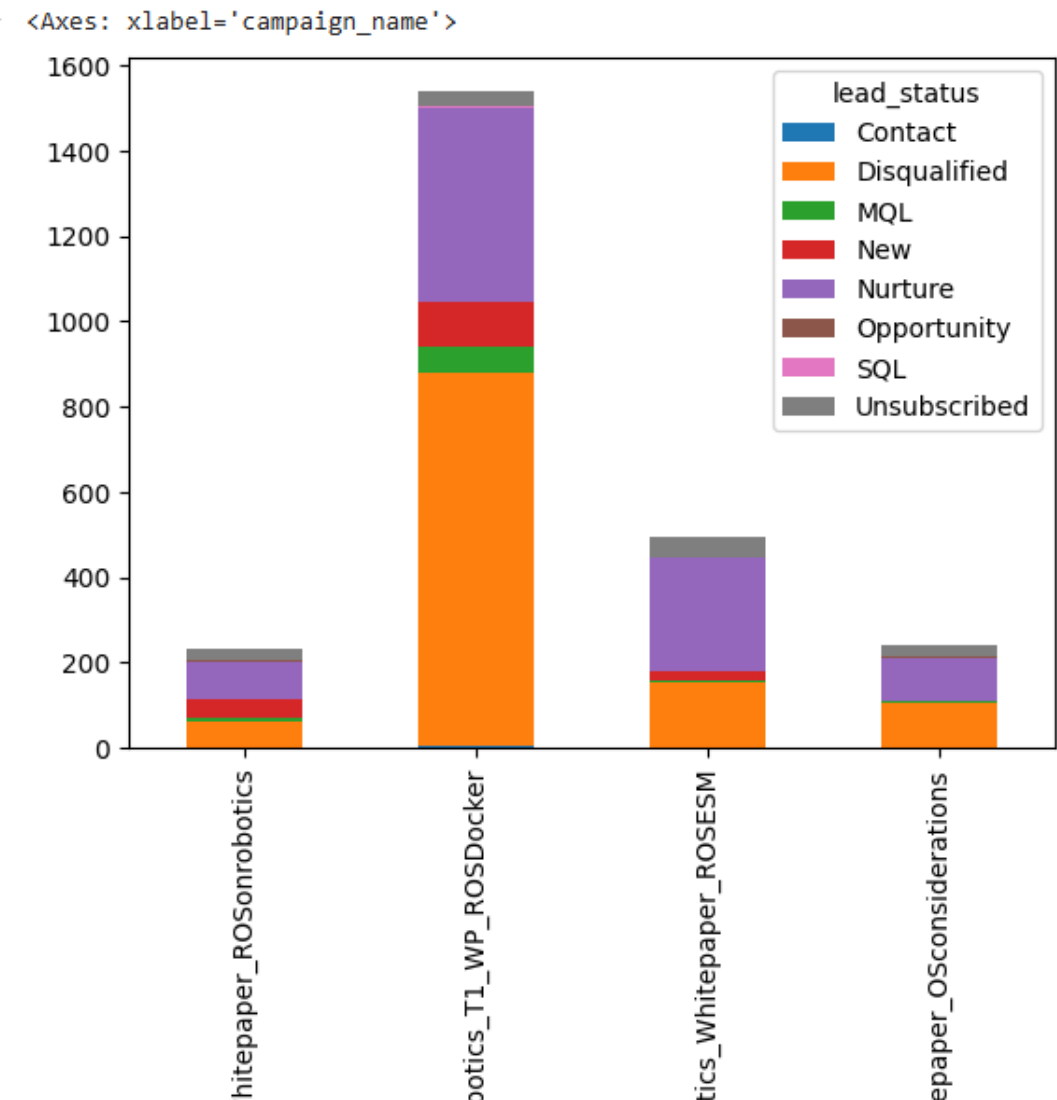
Engagement Metrics:

- **Page Views:** Total page views - c3 > c4 > c1 > c2. Total views/campaign ranges between 1610 - 37122.
- **Average Page View Time:** c2 > c3 > c4 > c1. Range is 1.8 mins -1.52 mins.
- **Bounce Rate:** c3 > c2 > c1 > c4. Range - 49.14% - 78.97%.

Lead Quality Metrics:

- **Conversion Rate:** % of convertible leads (only counting for - MQL, SQL, Opportunity) - highest secured by -> c3 > c2 > c1 > c4. Range - Min - 5, Max - 66.
- **Overall Lead status distribution:** Understanding the proportion of leads in different stages. Overall performance by campaign - c3 > c4 > c1 > c2.

```
#plot a stacked bar chart to show the breakdown of each lead_status for each campaign
marketing_campaigns_merged_df.groupby('campaign_name')['lead_status']\
.value_counts().unstack().plot(kind='bar', stacked=True)
```



```
#add a column called 'avg_page_views' as an average of total_page_views for each campaign in the aggregated dataframe
aggregated_page_data['page_avg_views'] = aggregated_page_data['total_page_views'] / 4
aggregated_page_data.head()
```

	campaign_name	total_page_views	page_avg_time	page_bounce_rate	view_date_range	page_avg_views
0	CY20_IOT_Robotics_Whitepaper_ROSonrobotics	1610.0	0 days 00:01:52.437288135	59.327898	2021-01-01 00:00:00 to 2021-12-31 00:00:00	402.5
1	CY21_IOT_Robotics_T1_WP_ROSDocker	37122.0	0 days 00:01:13.875000	78.965500	2021-11-22 00:00:00 to 2021-12-31 00:00:00	9280.5
2	CY21_IOT_Robotics_Whitepaper_ROSESM	9630.0	0 days 00:01:27.571428571	49.142925	2021-07-05 00:00:00 to 2021-12-27 00:00:00	2407.5
3	FY19_IOT_Robotics_Whitepaper_OSconsiderations	2048.0	0 days 00:01:08.165584415	51.219058	2021-01-01 00:00:00 to 2021-12-31 00:00:00	512.0

ALL LEAD STATUSES - DATA VISUALIZATION

C 3

>

C 4

>

C 1

>

C 2

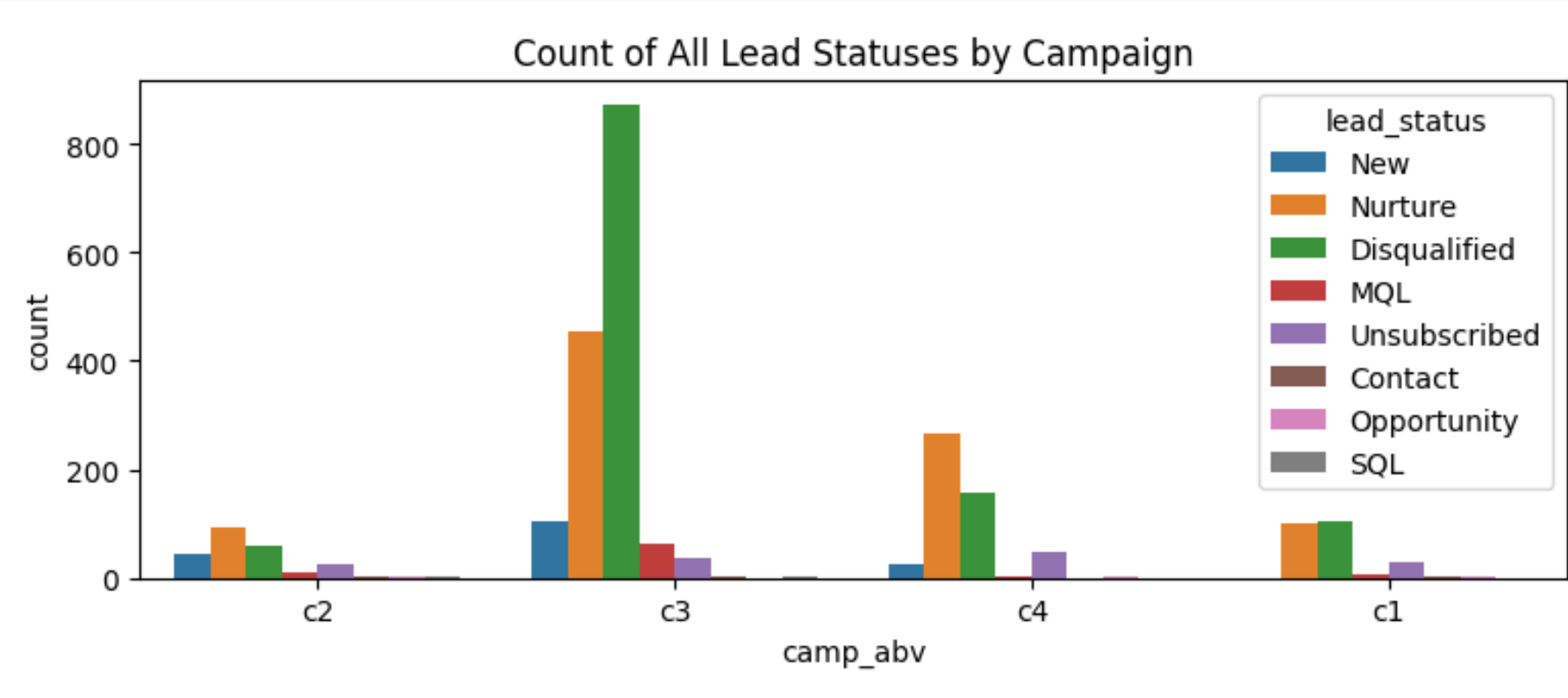
```
# FOR ALL LEADS

plt.figure(figsize = (15,12))
#count plot
plt.subplot(4,2,1)
# Use the column name 'campaign_name' instead of the 'variable' list
sns.countplot(x = 'camp_abv', hue = 'lead_status', data = df_cat)
plt.title('Count of All Lead Statuses by Campaign')

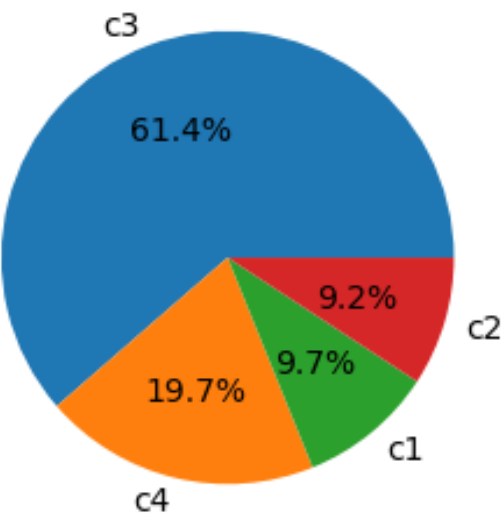
#Pie chart
plt.subplot(4,2,2)
counts = df_cat['camp_abv'].value_counts()
plt.pie(counts, labels = counts.index, autopct = '%1.1f%')
plt.title('Percentage of Lead Statuses per Campaign')

#Adjust layout
plt.tight_layout()

#Show the plots
plt.show()
```



Percentage of Lead Statuses per Campaign



‘CONVERTIBLE’ LEAD STATUSES - DATA VISUALIZATION

C 3

>

C 2

>

C 1

>

C 4

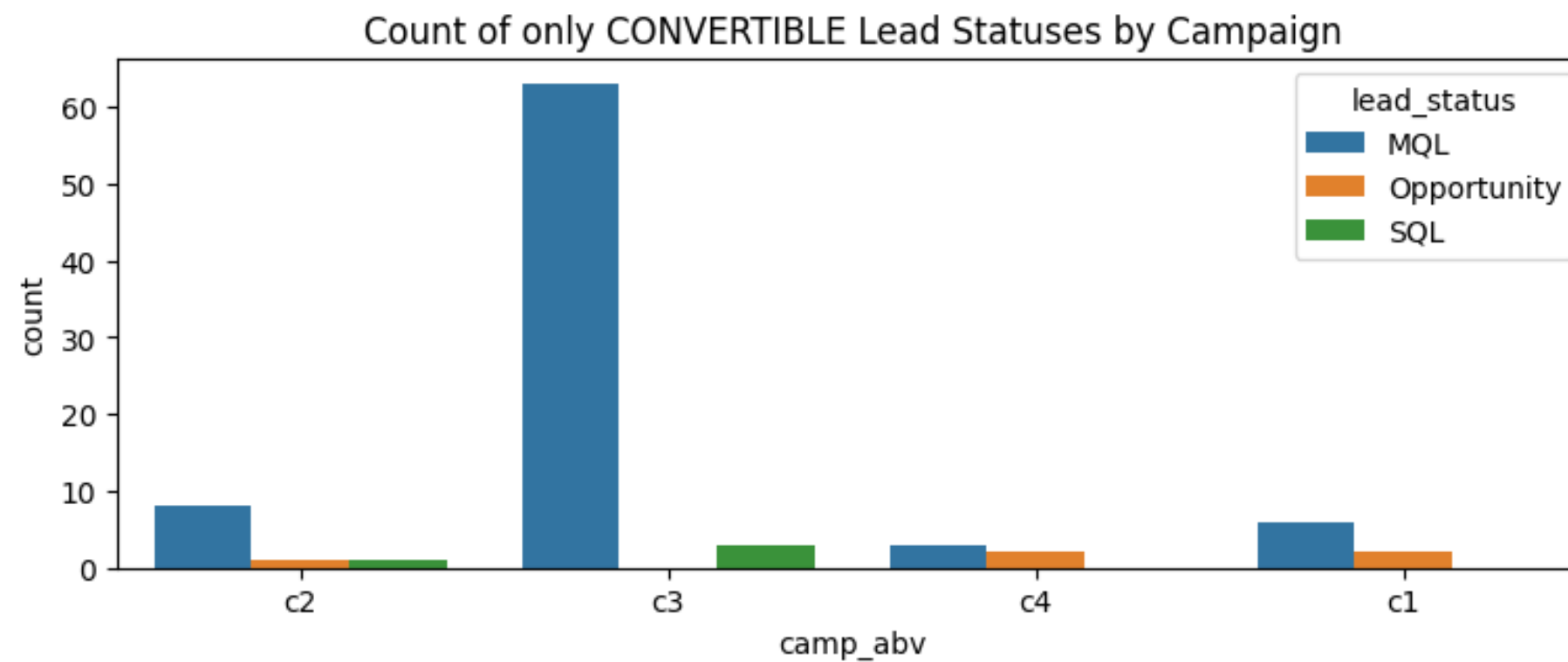
```
#Below only for Convertible leads (MQL, SQL Opportunity)

plt.figure(figsize = (15,12))
#count plot
plt.subplot(4,2,1)
# Use the column name 'campaign_name' instead of the 'variable' list
sns.countplot(x = 'camp_abv', hue = 'lead_status', data = df_cat_convertible)
plt.title('Count of only CONVERTIBLE Lead Statuses by Campaign')

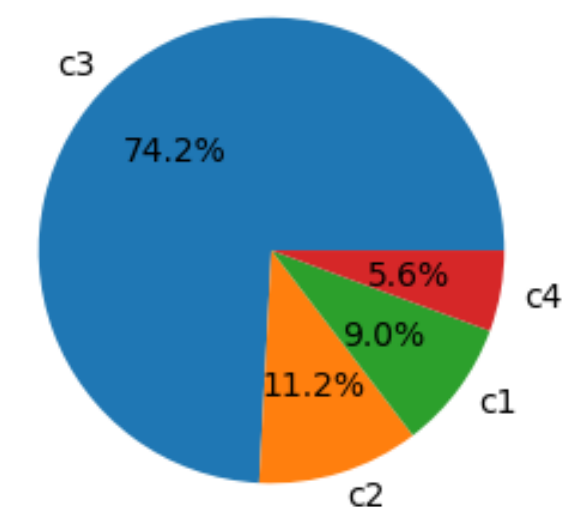
#Pie chart
plt.subplot(4,2,2)
counts = df_cat_convertible['camp_abv'].value_counts()
plt.pie(counts, labels = counts.index, autopct = '%1.1f%%')
plt.title('Distribution of CONVERTIBLE Lead Statuses')

#Adjust layout
plt.tight_layout()

#Show the plots
plt.show()
```



Distribution of CONVERTIBLE Lead Statuses



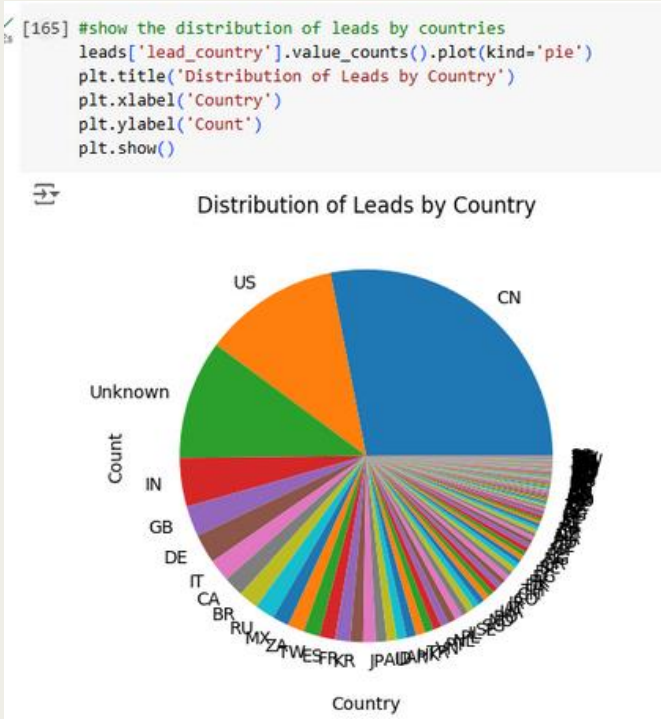
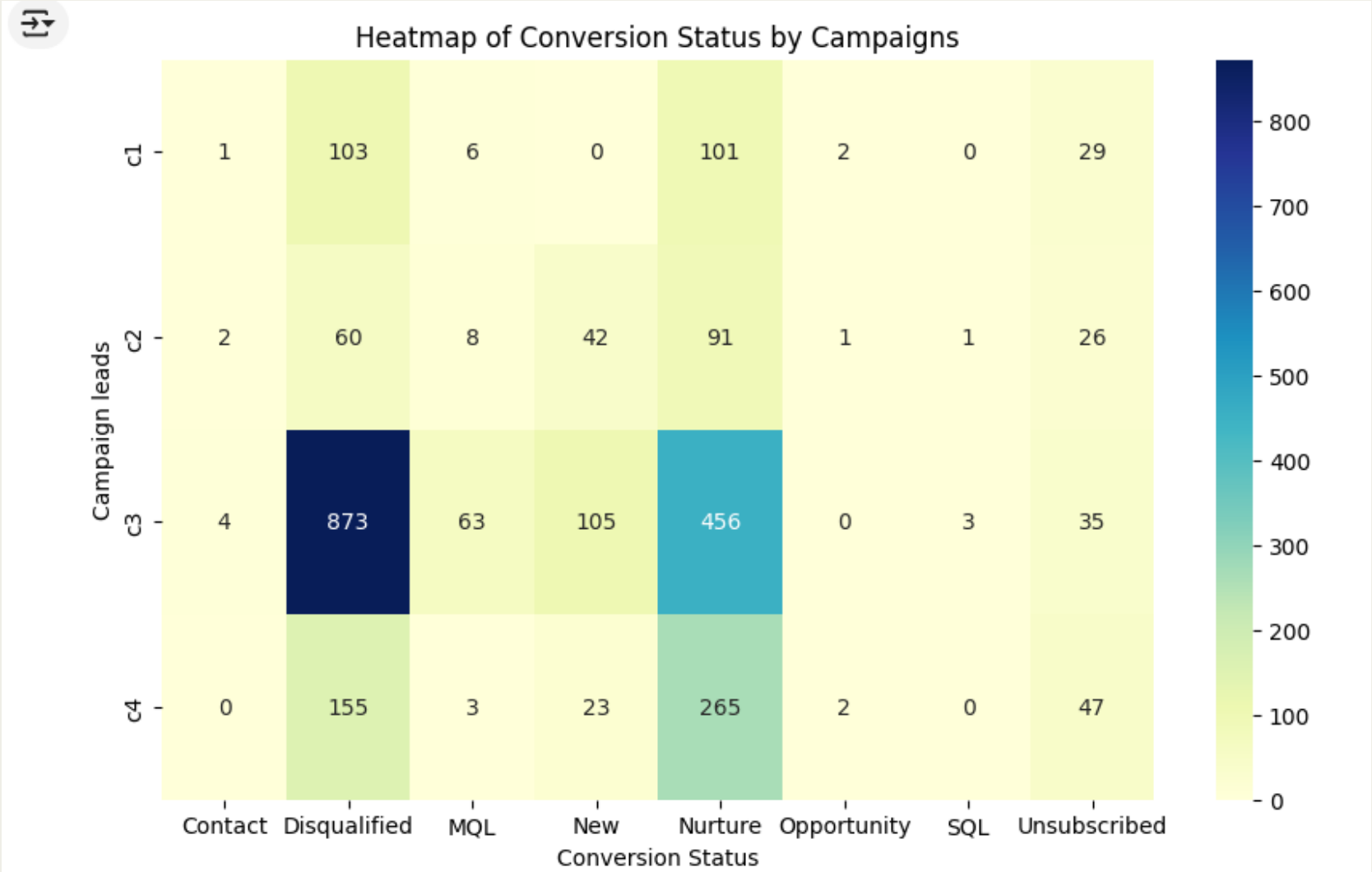
CAMPAIGN PERFORMANCE ANALYSIS - CONTD..

What's Working:

- **Distributed population:** Good dispersion in terms of global engagement comprising users from various countries and sources (Google, Facebook, Twitter, LinkedIn, etc).
- **High ‘Nurture’ leads:** high number of Nurture leads indicate user’s interest in moving to further stages in the market sales funnel thereby leading to possible conversions.
- **High Page views:** avg page views for each campaign is comparatively good (402-9280), given the dataset size, which can be tapped to lead to more possible conversions.

What's Not Working:

- **High Bounce rates:** high average bounce rate across all campaigns suggests lower user attention span on the portal. UI restructuring (button placement, catchy title, etc.) could help.
- **High ‘Disqualified’ leads:** very high no of lost leads under ‘Disqualified’ and ‘Unsubscribed’ categories, targeting these segments should be prioritized.
- **Hindered Source Effectiveness:** lots of high conversion leads have missing source information. Its crucial to identify marketing areas for highest-performing sources.
- **Low segmental optimization:** owing to the low average time spent on the pages, there’s not much customer-centric content on the campaign pages, which must be tailored.



CAMPAIGN PERFORMANCE ANALYSIS - CONTD..

Campaign 1: FY19_IOT_Robotics_Whitepaper_OSconsiderations

- Pros: 2nd least no of ‘Disqualified’ leads (155)
- Cons: 3rd in page views, bounce rate, and leads

Campaign 2: CY20_IOT_Robotics_Whitepaper_ROSonrobotics

- Pros: Least no of ‘Disqualified’ leads (60), 2nd in highest convertible leads (MQL, SQL, Opportunity)
- Cons: Least page views and overall leads

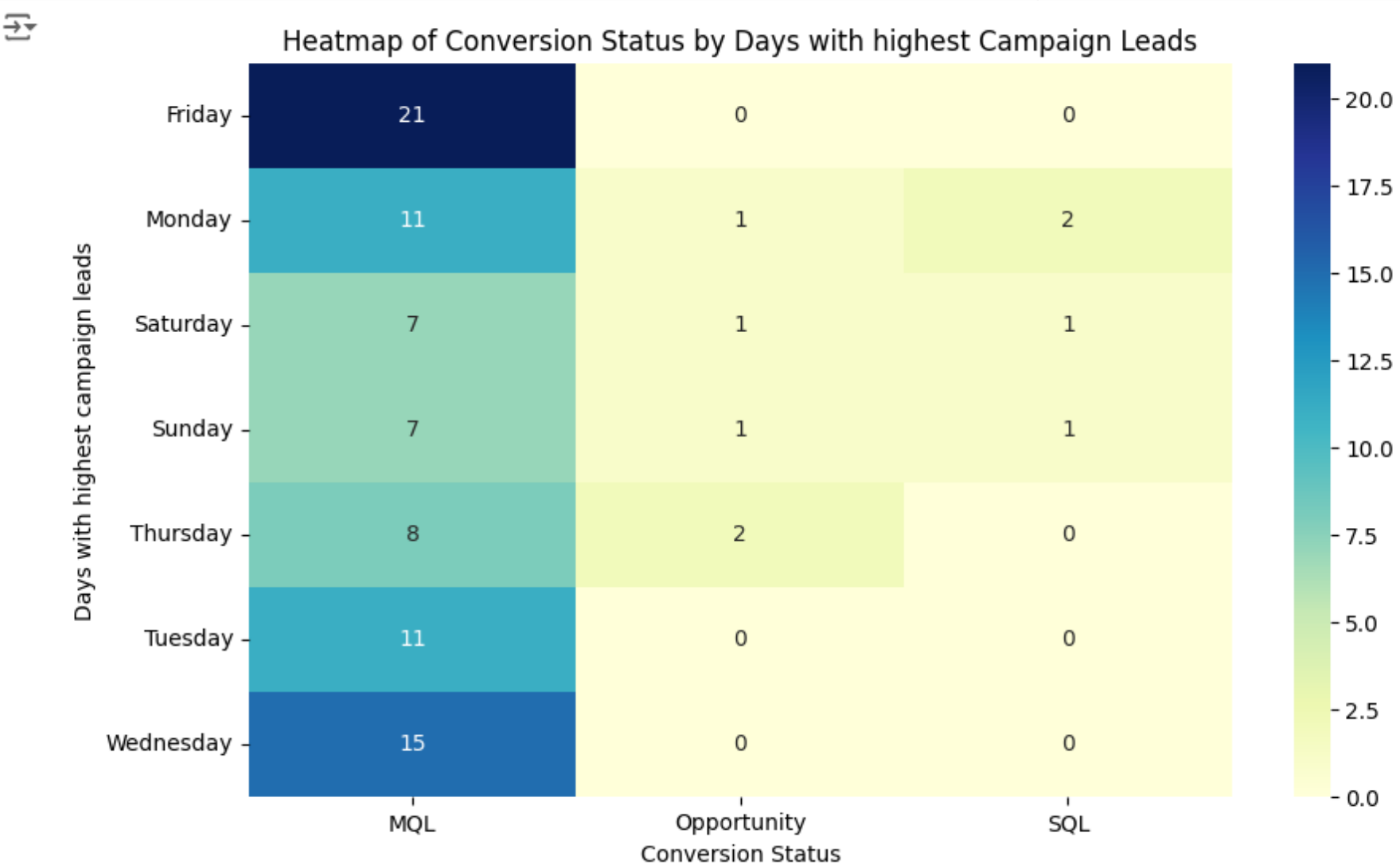
Campaign 3: CY21_IOT_Robotics_T1_WP_ROSDocker

- Pros: Best performer in page views (37122) and no of leads (1539)
- Cons: Highest no of ‘Disqualified’ leads

Campaign 4: CY21_IOT_Robotics_Whitepaper_ROSESM

- Pros: 2nd best campaign in page views (495) and no of leads (9630)
- Cons: Least no of ‘Convertible’ Leads (MQL, SQL, Opportunity)

```
# Visualize the contingency table with a heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(contingency_table_day, annot=True, cmap="YlGnBu", fmt='d')
plt.title('Heatmap of Conversion Status by Days with highest Campaign Leads')
plt.xlabel('Conversion Status')
plt.ylabel('Days with highest campaign leads')
plt.show()
```



BYTES OF INSIGHTS

Campaign Performance

#1 Campaign - C3



CY21_IOT_Robotics_T1_
WP_ROSDocker

C3 emerged as the most effective campaign in generating opportunistic and overall leads, with a good mix of paid, organic, and unknown lead sources.

#4 Campaign C1

Needs more work

Campaigns 1 & 2 – (FY19_IOT_Robotics_Whitepaper_Osconsiderations and CY20_IOT_Robotics_Whitepaper_ROSonrobotics) relatively fared similarly poorly throughout all the metrics of effectives - page views, leads, successful conversions, etc.

Seasonality and Exposure

Nov and Dec

Best months

November and December are top-performing months going by campaign joining dates and page viewing dates.
May, July, Sept, Aug, Oct are relatively slower months.

1500 - 2500 & 37000-37499

Optimal page views

The ‘sweet spot’ ranges for optimal number of pageviews for a campaign to secure conversions is 1500-2500 and 37000-37499.

Hot/Cold Days

Mon, Wed, Weekend

For better converts

Weekends fared better for sales leads, and Monday, Wed, Fri performed better for Opportunities and Marketing Qualified Leads.

Thursday, Tuesday

Highest lost leads

Thursday and Tuesday saw the highest number of ‘Disqualified’ and ‘Unsubscribed’ leads.

Social Demographics

Student, Engineer

Top titles for leads

Top professions which have highest leads for ROS in the dataset.

CN, US, IN

Top lead countries

These countries make the top 4 countries for overall leads, 3rd being the ‘Unknown’ category, making up for the missing data.

RECOMMENDATIONS

Improve data quality

- Streamline data collection pipelines to capture qualitative data for critical fields (lead sources, country, industry, etc.)
- Improve industry specific data to target domain-specific demand for ROS.

Content optimization

- Based on the average time spent on pages, adjust the content to enhance segmented user engagement.
- For pages with high bounce rates and low time on page, consider revising the content to better match user expectations.

Source effectiveness

- Evaluate lead_source to identify which channels are driving the most conversions.
- Focus marketing spend on the highest-performing sources.

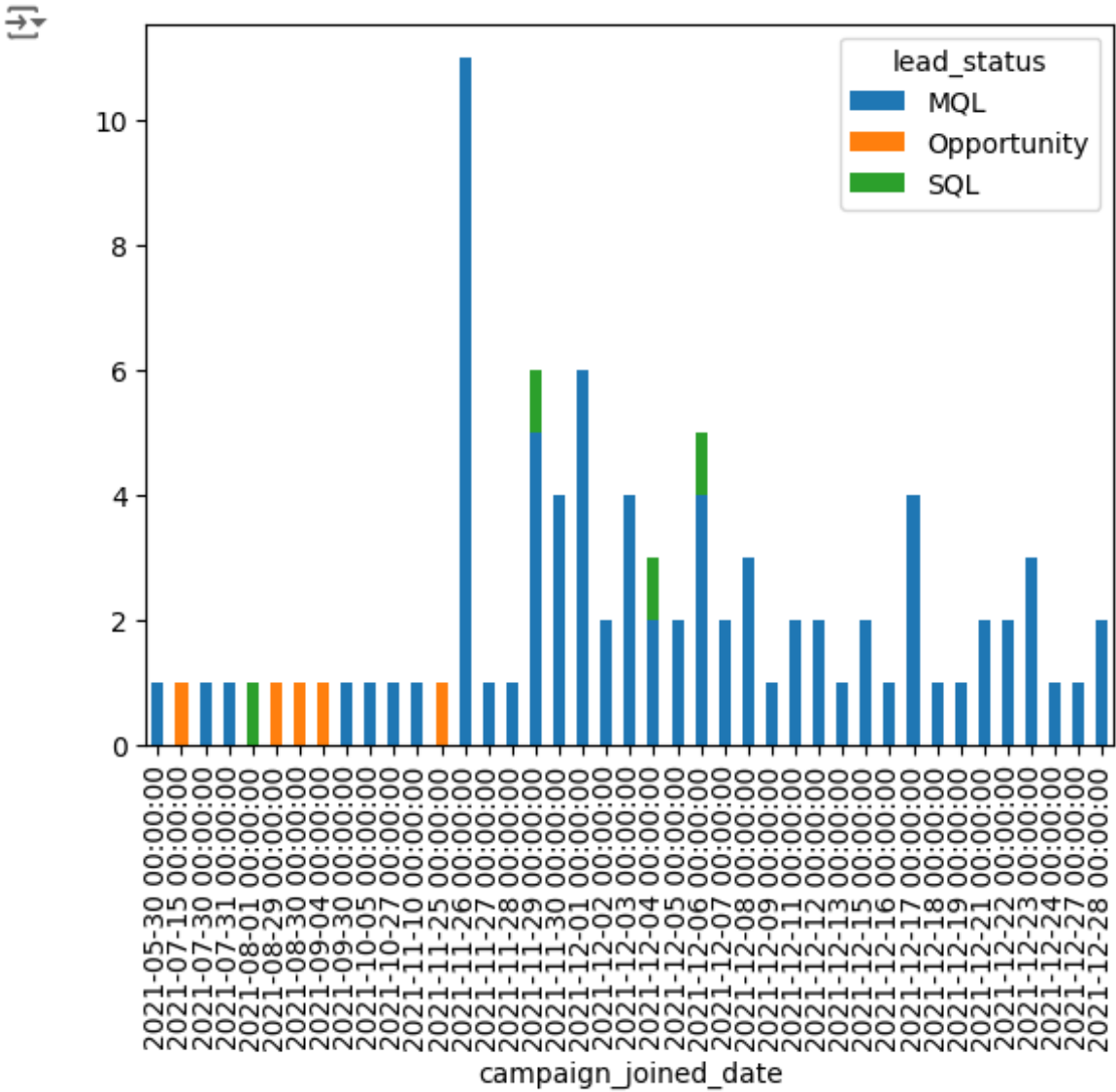
Leverage Seasonality trends

- Customized content centered on seasonality theme may attract conversions.
- Consider allocating more marketing budget to months with higher conversions and engagement (Nov, Dec in this case) to maximize ROI.
- Structure subscription offers and discounts for festive months to get people to spend more time on the portal.

We can perform a Cohort Analysis, by grouping leads by `campaign_joined_date` to assess campaign performance.

Below is to understand on which campaign joining days the 4 campaigns saw the highest conversions or best possible leads.

```
campaign_conversion_day = pd.crosstab(conversions_df['campaign_joined_date'], conversions_df['lead_status'])
campaign_conversion_day.sort_values(by = 'SQL', ascending = False)
campaign_conversion_day.plot.bar(stacked = True);
```



FUTURE DATA ANALYSIS STRATEGIES

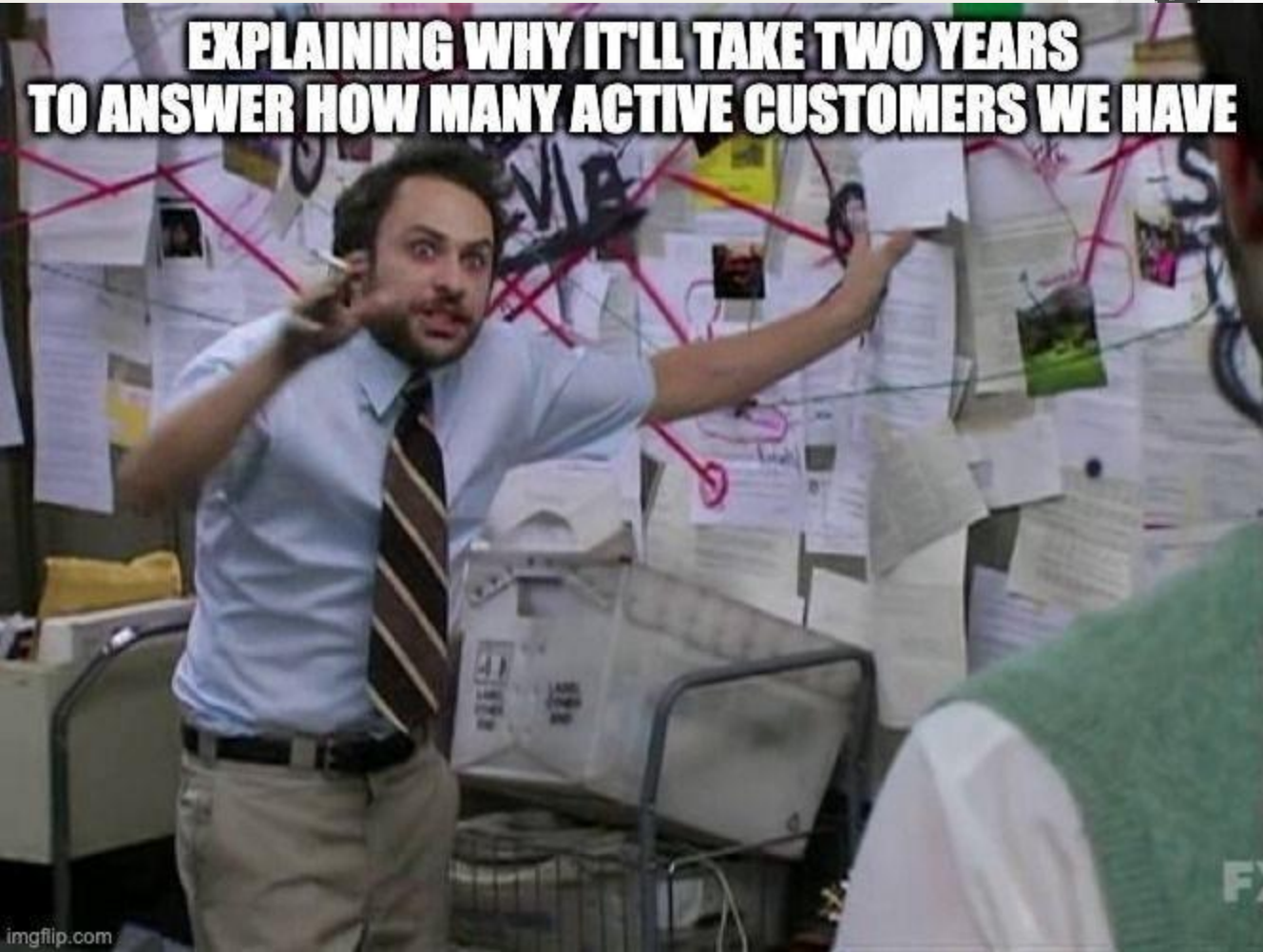
Additional Data Needed:

Customer data:

- **Lead Interaction Data:** Details on how leads interact with your brand.
- **Lead Score:** A calculated value based on various factors indicating lead quality.
- **Customer Feedback:** Gather direct insights from customers to improve your offerings.

Revenue Data collection:

- **Campaign Cost:** gather pricing related data (e.g., ad spend) to measure the effectiveness of the campaigns to track expenses.
- **Customer Lifetime Value (CLTV):** Estimate the total revenue a customer will generate over their lifetime.



```
#missing value pecentage in Leads
(leads.isna().sum()/len(leads))*100)\
sort_values(ascending=False)\
reset_index(name='Missing Values %')
```

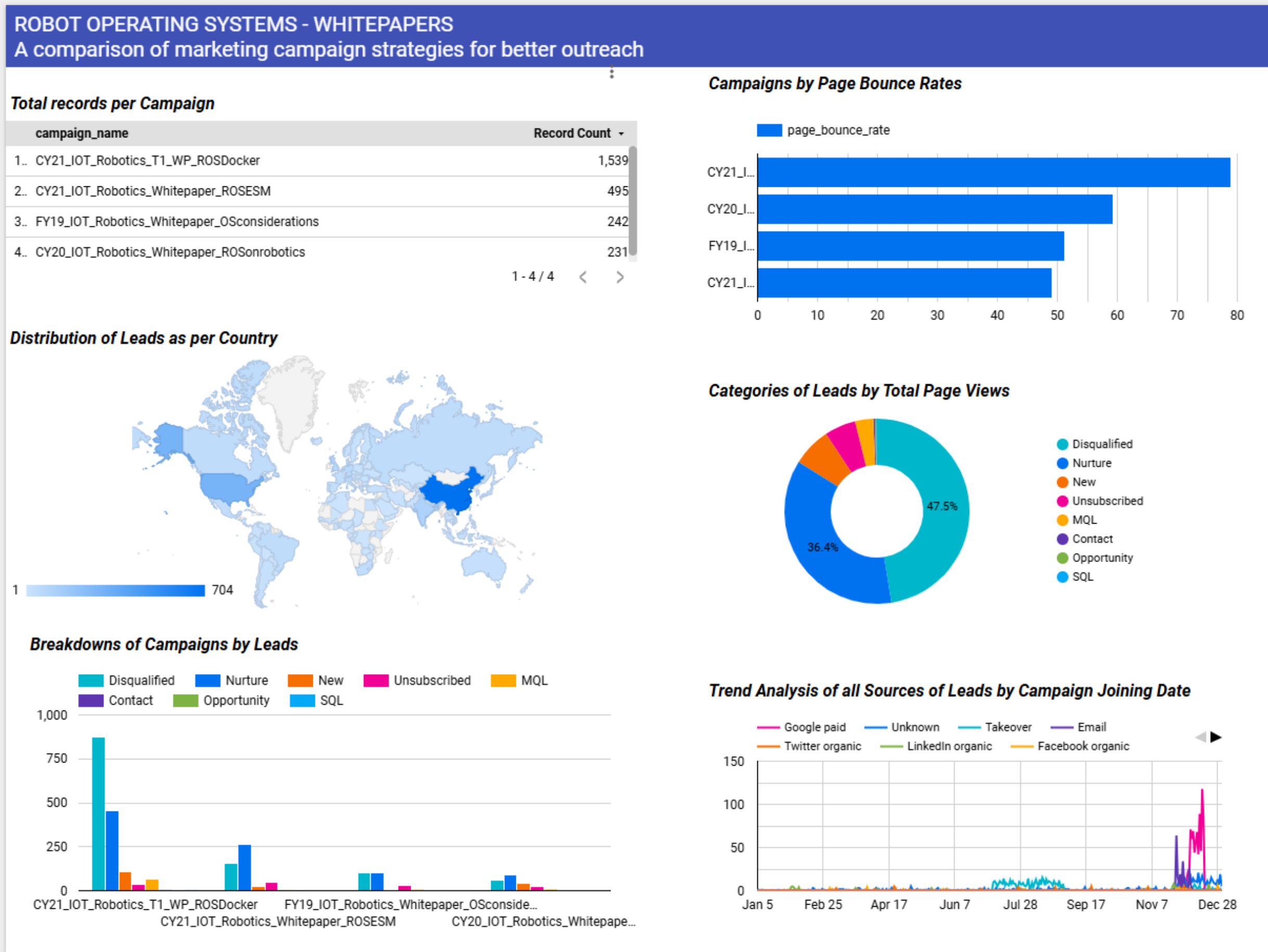
	index	Missing Values %
0	lead_industry	84.488189
1	lead_source	25.196850
2	lead_country	10.354331
3	lead_job_title	1.299213
4	lead_hashed_id	0.000000
5	campaign_joined_date	0.000000
6	campaign_name	0.000000
7	lead_status	0.000000



CONCLUSION

●	●	●	●	●
<i>Analyse</i> data	<i>Compare</i> baseline	<i>Quantify</i> metrics	<i>Monitor</i> market	<i>Commercialize</i> revenue
Began with analysis of all datasets and chose appropriate ways of handling and cleaning to obtain a final dataset with all quantifiable columns for improved analysis.	Comparison on basis of baseline metrics such as leads, page views, seasonality trends, etc.	Undertake cross-testing and inferential statistical analysis to obtain actually significant trackers of performance analysis for various marketing campaigns.	Understand the blockers, create A/B tests over a period of time to have an evidence-backed strategy planning about which campaigns need help and which are actually converting to revenue.	Work on suggested data strategies to focus on areas like profit margin, cost-per-campaign, outreach mediums, etc., to increase revenue per campaign and savings.

By focusing on both engagement and lead quality metrics, we have determined what’s working and what’s not, and now need to implement data-driven changes for future campaigns.



Join us in revolutionizing access to
‘Robotic’ ideas and emerging
technologies.

Thank you!

