Data Analysis of Internshala

PREREQUISITE:

Imagine you are working for an organization that offers advanced certifications in various courses.

OBJECTIVE:

To analyze a dataset that contains information about how leads are acquired, categorized and converted, and also highlighting trends, patterns, or any other meaningful observations that could help the organization understand how leads are generated and how they interact with courses.

Data Source:

link

By
Sujal Dubey

INTERNSHALA

Data Summary

Total Leads

16,460

Total Leads Paid

648

Total Channel Groups

13

Total Courses Offered

7

TOP SUBSTANTIAL INSIGHTS

1) **Channels performance:**
   a) Out of 13 Channel groups **A** is the most effective for lead generation and conversion, producing **7,932 leads** alone and **313** successful payments.
   b) Channels **B, C** and **D** have the **highest conversion rates**, combined generating over **1 Crore** in amount paid, though with significantly lower in count.

2) **Course Interaction:**
   a) **Python and Java** are the most popular courses among leads, bringing more than **50% of Total Leads.**

3) **Payment Trends:**
   a) Out of 16,460 leads, only **648 paid**, with **Java** and **Python** being the most purchased courses(**313**).
   b) Leads generated through **interacting with EFG** gripped **over 50%** of the total payments.

4) **Conversion Ratios:**
   a) Channel **B** has a conversion rate of **over 15%** for leads turning into paid leads.

5) As per the data, **on average**, **leads purchase** the course **within 8 to 10 days** from the 'Lead Date'.(Min. 1 day or Max. 28 days).(Assumption 4.)

6) Peak of **569 leads** was recorded by the Graduates of **2024** throughout graduation years.

ANALYZING HOW LEADS ARE GENERATED, CATEGORIZED AND CONVERTED

live dashboard

INTERNSHALA DATA ANALYSIS

Select Course: All

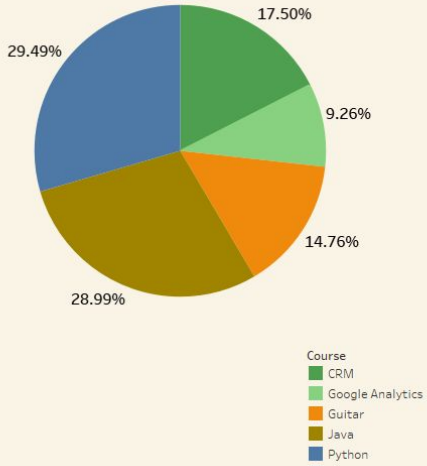
Select Channel group: All

TOTAL LEADS	TOTAL LEADS PAID	TOTAL CHANNEL GROUPS	TOTAL COURSES OFFERED	Highest Interaction Type
16,460	648	13	7	ABC 6,652

Note Worthy Visualization (click visuals to filter data)

Top 5 Channels for generating leads:

Top 5 Courses showing interest



Leads Per Year

Probability of Leads Converting

	A	B	C	D	E	F	G	H	I	J	K	L	M
CRM	6.45	15.18	11.49	10.88	3.76	0.26	1.27	3.09	0.00	0.00	0.00	0.00	0.96
Figma	3.22	15.56	8.33	3.85	1.39		0.00	1.59	0.00		0.00	0.00	0.00
Google Analytics	2.66	17.91	5.66	5.83	0.00		0.00	2.74	0.00		0.00	0.00	0.44
Guitar	4.08	24.07	4.65	10.40	0.00	0.95	2.56	2.78	0.00	0.00	0.00		1.25
Java	3.63	18.64	11.18	6.65	1.73	0.19	1.72	2.20	0.00	0.00	0.00	0.00	1.19
Python	3.45	13.10	8.38	6.17	1.93	0.00	1.19	1.37	1.19		0.00	0.00	0.41
SEO	4.48	18.18	2.08	2.70	2.35		0.00	2.33	0.00		0.00	0.00	0.00

Significant More Findings and Recommendations

1) **Channels performance:**
   a) **Python Course** should be included in **Channel Group J,** as it has proven highest leads generation.(Assumption 3)
   b) Introducing **“ABC” lead type in Channel B** will help boost in both conversions and generation of new leads.
   c) Using **“LMN”** lead type in high acquisition channels(**A & M**) will help in finding the effectiveness of the lead type.

2) **Course Preferences:**
   a) **Both Guitar** and **Google Analytics** also being popular but show lower conversion rates.
   b) Promote **Python, Java, and CRM courses**, as they show high lead interest and conversion potential.

3) **Payment Trends:**
   a) **Only 3.9%** of overall leads purchased the courses.
   b) Capitalize on 2024 lead generation spike with targeted campaigns and offers.(Assumption 2.)
   c) **CRM** has the **highest conversion ratio** of **5%** w.r.t. Leads acquired.

4) Prioritize Channel **A** for **both lead generation and conversion.**

5) If we had **dates** for when the **channel_group is created**, we can compare the performances of the groups.(Assumption 1.)

6) Deeper Visualizations for further data is done, probabilities of conversions, top 5 channels and courses, peak leads throughout graduation years.(Appendix 3)

Assumptions made:

1. Assuming, Data is available for when the “channel_group” is created.
2. Assuming, May-2024 had the highest leads recorded throughout the year 2024.
3. Assuming, Every course is offered in every channel.
4. Assuming, Availability of data for “paid_at” is from May,1st to June, 1st.

Appendix:

1. To see the 1st raw analysis in Microsoft Excel. click here
2. Insights are also generated and analyzed in SQL language using PostgreSQL, here is the link to queries used.
3. Also included deeper insights in the form of visualizations using Tableau, I have also attached link for the same.
4. Combined link to Google Drive is here.