Dataset Overview

Purpose of the Dataset: This dataset provides insights into how leads interact with the courses, including their origins, preferences, and payment behaviours. This will help identify trends, optimize lead generation, and improve course offerings.

Missing Values: The amount_paid and paid_at columns have missing values. *These missing values probably indicate leads who viewed the courses but did not buy.*

Total Record: The dataset consists of 16,460 records**.**

Top 5 Insights from the Dataset

1. **Channel Group Distribution: Identify the number of leads generated per channel group.**

SQL Query:

```
6 •  SELECT
7        Channel_group,
8        COUNT(lead_id) AS Lead_Count
9    FROM tabletest
10   GROUP BY Channel_group
11   ORDER BY Lead_Count DESC;
```

100% 26:11

Result Grid | Filter Rows: Search | Export:

Channel_group	Lead_Count
A	7932
M	1647
F	1586
D	1294
E	1080
H	785
B	741
C	671
G	439
I	211
K	34
L	28
J	12

Result 15

Fig: Screenshot from MySQL Workbench (query and output)

Graph Representation:

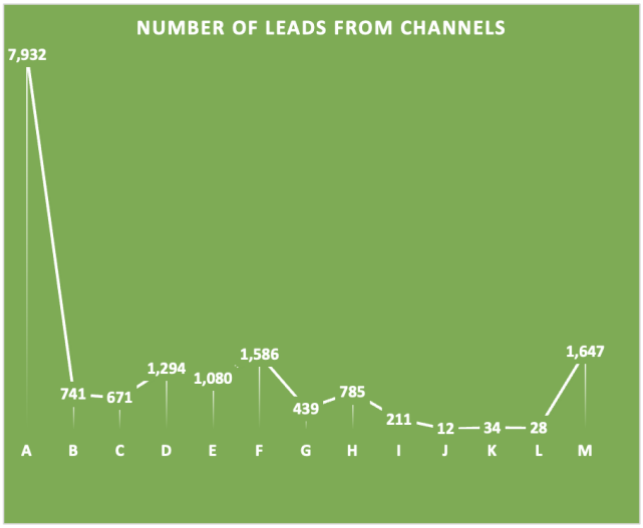

Fig: Line graph representation of channel_group(x-axis) vs lead_id(y-axis)

Analysis: Most leads originate from channel *group A, 7932* leads, then *M, 1647* leads, and *F, 1586* leads. J and L have fewer leads, so these channels could be underperforming.

Assumptions:

- Channels M (1,647 leads) and F (1,586 leads) perform moderately compared to A. The channels may be optimized or have more investment in them to achieve better performance.
- Channels J and L have significantly fewer leads. These channels seem to be underperforming. These channels might Require revised strategies to increase effectiveness.

2. **Most Popular Courses: Finding the most popular courses based on the number of leads.**

SQL Query:

```
12
13 •  SELECT
14        course,
15        COUNT(lead_id) AS Lead_Count
16    FROM tabletest
17    GROUP BY course
18    ORDER BY Lead_Count DESC;
```

100% 26:18

Result Grid | Filter Rows: Search

course	Lead_Count
Python	4323
Java	4250
CRM	2565
Guitar	2164
Google Analytics	1358
Figma	975
SEO	825

Fig: Screenshot from MySQL Workbench (query and output)

Graph Representation:

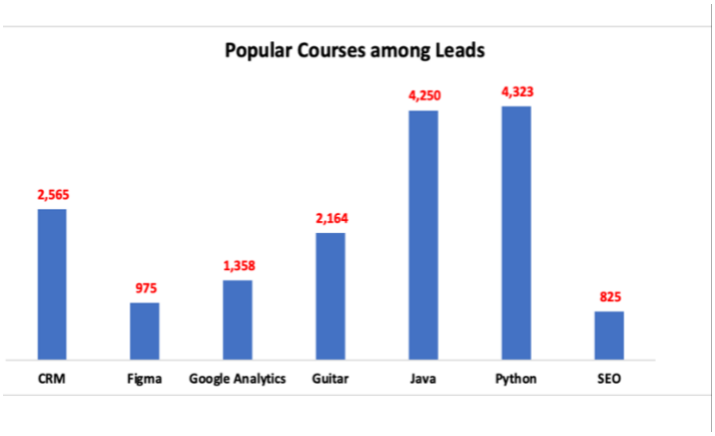

Fig: Bar graph representation course (x-axis) vs lead_id(y-axis)

Analysis: There is a huge preference for the *Python course with 4,323 leads*, closely followed by the *Java course at 4,250 leads*. The *CRM and Guitar* courses follow suit with *leads above 2,000 leads*, while leads *in SEO and Figma courses are fewer in number.*

Assumptions:

- The Python course is probably the most in-demand because it applies to the most in-demand fields, including data science, machine learning, and software development.
- These might indicate the presence of some potential in the market for SEO and Figma skills if such skills are demanded but under-marketed. Conversely, these courses could be needed in terms of their content or structure to generate more leads.

3. **Lead Type Distribution: Analyze lead types and their payment details.**

SQL Query:

```
22 •  SELECT
23        Lead_type,
24        COUNT(lead_id) AS Total_Leads,
25        SUM(amount_paid) AS Total_Amount_Paid
26        COUNT(DISTINCT paid_at) AS Paid_Leads
27    FROM tabletest
28    GROUP BY Lead_type
29    ORDER BY Total_Amount_Paid DESC;
```

100% 33:29

Result Grid Filter Rows: Search Export:

Lead_type	Total_Leads	Total_Amount_P...	Paid_Leads
EFG	6218	15278810	361
ABC	6652	10562791	255
XYZ	3260	1160765	29
LMN	330	0	1

Fig: Screenshot from MySQL Workbench (query and output)

Graph Representation:

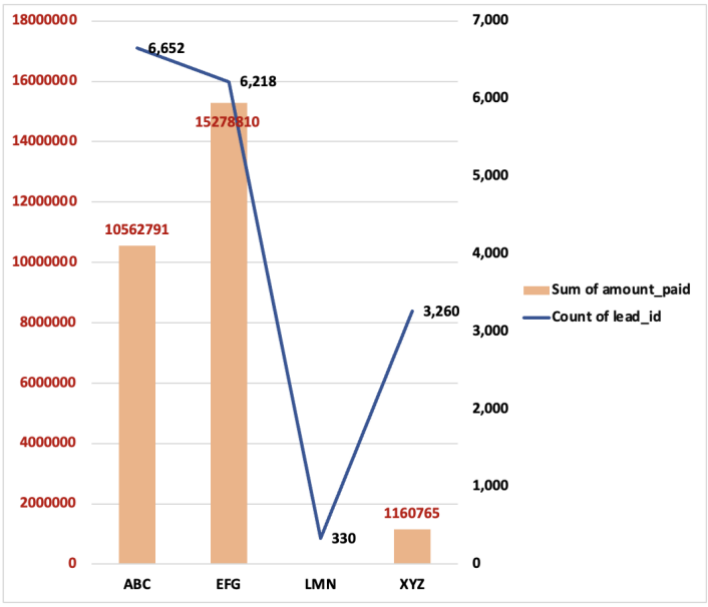

Fig: Combo chart representation lead_type (x-axis) vs lead_id and amount_paid(y-axis)

Analysis:

- Lead type ABC constitutes the largest share (6,652 leads), followed by EFG (6,218 leads). XYZ leads are fewer (3,260), and LMN is minimal with only 330 leads.
- While ABC has the highest total leads (6,652), its paid leads are fewer than EFG, which has fewer total leads but higher paid conversions. This suggests EFG is more efficient at converting leads into paying customers.
- XYZ and LMN have significantly lower total leads and paid leads, indicating these lead types are either less targeted or less effective in generating paying customers.

Assumptions:

- EFG is the best lead type in producing revenue, as though EFG had fewer leads than ABC, this shows that its campaigns or strategies are more conversion-oriented.
- ABC may have many leads but conversion of leads into a paid customer basis may not be so great.
- LMN doesn't seem effective in revenue generation, which would either indicate these leads are not targeted for paid conversions or the lead type simply reflects a non-paying audience.
- XYZ and LMN may need focused efforts to increase conversion rates or be reassessed if their poor performance continues.

4. **Lead Generation Trends: Count the number of leads generated in each month.**

SQL Query:

```
30
31 •  SELECT
32        EXTRACT(MONTH FROM lead_date) AS Month,
33        COUNT(lead_id) AS Lead_Count
34    FROM tabletest
35    GROUP BY EXTRACT(MONTH FROM lead_date)
36    ORDER BY Month;
```

100% 19:28

Result Grid | Filter Rows: Search | Export:

Month	Lead_Count
5	16460

Fig: Screenshot from MySQL Workbench (query and output

Analysis: All leads were generated in May, indicating this may be a snapshot of monthly data or a seasonal pattern that requires further investigation.

Assumptions: May could be a peak month for lead generation, driven by marketing campaigns, course launches, or seasonal demand.

5. Graduation Year Distribution: Distribution of leads by graduation year.

SQL Query:

```
28
29 •  SELECT
30        graduation_year,
31        COUNT(lead_id) AS Lead_Count
32    FROM tabletest
33    GROUP BY graduation_year
34    ORDER BY graduation_year;
```

100% 8:29

Result Grid Filter Rows: Search Export:

graduation_ye...	Lead_Count
2000	556
2001	534
2002	565
2003	544
2004	550
2005	540
2006	494
2007	512
2008	524
2009	538
2010	504
2011	558
2012	486
2013	519
2014	506
2015	509
2016	543
2017	542
2018	519
2019	545
2020	551
2021	520
2022	504
2023	512
2024	569
2025	505
2026	546
2027	553
2028	558
2029	526
2030	528

Result 23

Fig(left): Screenshot from MySQL Workbench (query and output)

Analysis: The dataset covers leads who graduated between 2000 and 2030. The most leads are found in *the years 2024 (569 leads) and 2028 (558 leads)*, indicating that there is a greater interest from people who are closer to graduating.

Assumptions:

- This could be because leads who are either soon graduating or graduated with a degree are more interested in pursuing the courses offered. It could also be because of career preparation, an upgrade of skill, and professional development needs.
- Targeted Marketing will be done, focusing on the graduating class, 2024 and 2028, using tailored campaigns for career-related benefits.

Graph Representation:

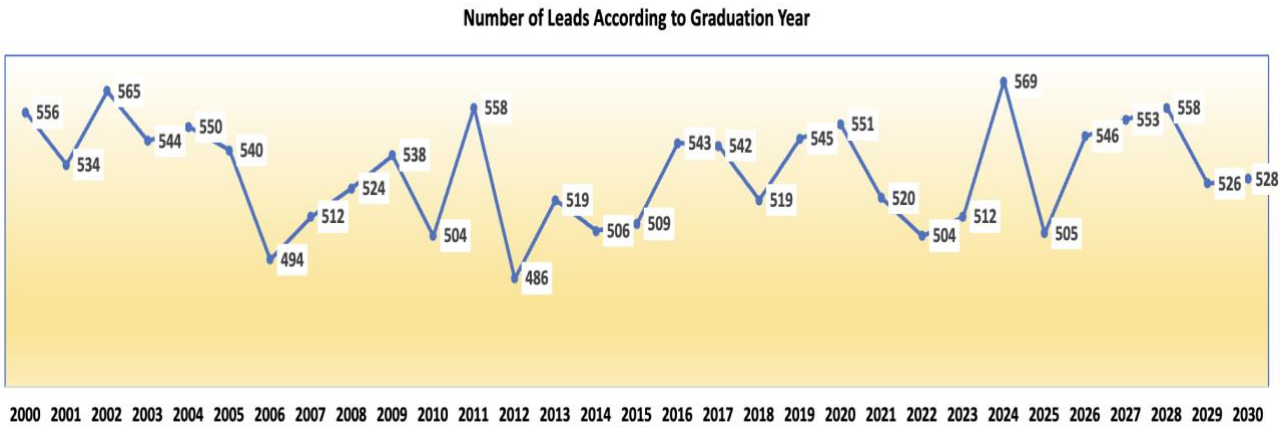

Fig: Line graph representation of graduation-year (x-axis) vs lead_id(y-axis)