Capstone Project

Analyzing Educative Courses Data

Introduction

You are a Data Analyst working for the education tech company, Educative. You have been asked by your manager, Head of Curriculum at Educative, to present the data on course revenue, and you have been provided with data on courses from different topics to understand where opportunities to increase revenue may lie and track the performance of courses.

Your manager has suggested encouraging Web Development courses to charge more, because she believes that these are the most popular courses. She needs to send a report to the CEO in the next three weeks on how they will increase their next quarter earnings.

Main tasks:

- Analyze the data from the company in line with Cross Industry Standard for Data Mining (CRISP-DM) and write the report for your manager.
- Create an interactive dashboard in Excel showing your visuals and metrics.
- The CEO is a lover of Power BI hence requires an interactive Dashboard with Power BI in addition to Excel the dashboard.

Further information

The information below is to help provide guidance to succeed in this project but you are not limited to it.

Data Analysis

1. Framing the Problem

Analytic problem framing involves translating the business problem into terms that can be addressed analytically with data and modeling.

When framing a problem, remember to:

- Get to the root of the business problem (sometimes the business talks solutions instead of problems)
- Get a gauge of what the timelines are supposed to be and how much funding there is (if possible).
- Understand where the data sits.

Task

Answer the following:

- What is the business problem?
- How long do you have to work on this project?
- Who are the stakeholders of this project?
- What data should be collected to understand this problem? How should it be presented?
- What questions would you ask to better understand the business problem?

2. Collecting and Cleaning Data

2.1 Collection

Data comes from multiple sources, and data analysts often gather data from multiple sources and combine them for data analysis. The process of gathering data from these sources and presenting it is called Data Consolidation or Data Collection.

Data Consolidation is a crucial step, as the accuracy of the insights from your data analysis depends heavily on the quality of data used.

Unzip the attached dataset into a folder. You will use the CSV files for your analysis.

Task

Import the data sheets (CSV) and consolidate the data in MS Excel.

Instructions

- Open Excel
- Go to the Data tab and click on the Get File dropdown. Follow this path to import the folder containing the data sheets: **Files > From Folder**
- Select the folder where the files are stored and click on **Open**.
- In the preview windows, click on the **Combine** button and select **Combine & Load**.

Now your data has been consolidated into one Excel sheet file. You can go through the data to view its content.

2.2 Data Cleaning

Good data is essential when using data to derive insights and make business decisions.

Data cleaning plays an important role in the analytical process and making sure that the answers we uncover are reliable and of a high quality.

You will use a few functions to clean our data in excel.

Task

Use the listed actions below to clean your data.

1) Remove duplicates

Select the entire datasheet to remove duplicates from and go to **Table Design > Remove Duplicates**

2) Remove blank cells

- On the '*subject'* column, click on the Filter icon, then deselect all except *(Blanks)*
- Blank Cells will arise to the top of the sheet. Remove these rows.

3) Proper Headers

Ensure you have clear and concise names for headers and use dashes or underscores in between words to make it easier to parse later.

4) Find and Replace

If you examine the data, you will see that the *Web Development* subject title is not the same as other subject titles. Use the Find and Replace function to make the Web Development subject consistent with other subjects. **Use: CTRL + F > Replace**

3. Data Analysis

With your cleaned Excel sheet, we're going to perform some functions to get additional data on the Educative courses and consolidate and present this clearly. To do this, you will be using the **Right/Left Function, IF function, VLOOKUP** and we will create Pivot Tables and Charts.

3.1 Spreadsheet Functions

Functions are used in formulas to perform specific actions. You will use functions to transform and analyze the data.

Task

1) While there is a published time available for each course, we want to know the date that each course was published without the time stamp.

Create a *'published_date'* column.

2) We want to know how many courses are free and how many are paid.

Create a column called 'free_or_paid' that indicates whether a course is Free or Paid.

3). Your manager has asked for a list of the top 20 most subscribed courses.

She has asked for

- Their level
- Whether they are free or paid
- Duration of the course
- The date they were published

Find and organize this information on a separate sheet. Your manager wants to understand the trends for the most subscribed courses.

First, sort the data in descending order based on *'num_subscribers'* column and copy the top 20 courses into the new sheet before using VLOOKUP to get the rest of the data.

3.2 Pivot Tables and Charts

Pivot tables are used for summarizing data. And charts are used for creating charts to visualize data. You will use both tools to analyze and gain insights from the data.

Task

Create the following Pivot Tables and Charts:

- Total number of subscribers for each subject (Pie Chart)
- Average subscriber count per subject (Bar Chart)
- Average cost per subject at each level (Bar Chart)
- Average content duration per subject (Bar Chart)
- Average rating per subject for each level (Column Chart)
- Any other information that you feel will be important to include in your analysis

You can place all pivot tables in one worksheet and all charts in a separate worksheet. Otherwise, you can keep both in the same worksheet but arranged properly.

Feel free to decide on what brings out your best work.

4. Understanding the Problem

Once you have framed the problem and gathered initial insights from the data, you can ask the following questions as you dig deeper into your analysis.

- What do you see happening?
- What are the specific symptoms of the problem?
- What is your hypothesis for the cause of the problem?

Task

Begin your analysis of the problem with your findings from the data.
What are your hypotheses and possible recommendations?

5. Visualizing your Data

Once done with your analysis, you will have to create an interactive dashboard you will use for telling the story of your analysis. You will use Power BI to create reports and visualization for your manager.

Task

Create reports with visualizations to support your analysis.

- Create a dashboard with all the relevant visualizations for your analysis in Excel.
- Create a dashboard with all the relevant visualizations for your analysis in Power BI.

Data Analysis Reporting

A good data story is a key vehicle to convey insights, with visualizations and data being important proof points. Facts simply present data; a story provides context, which augments our understanding and drives valuable insights. Creating a data story allows users to build trust in your analysis because you are answering their questions as they ask them and proving your insights to them.

A good data story leverages three major components - Data, Narrative and Visuals.

Data - requires having accurate data to reach correct insights.

Visuals - helps us comprehend the vast amounts of data collected. They provide at-a-glance snapshots of data, lacking the context needed to explain why something has happened.

Narrative - The third and most vital part of a data story. A narrative uses language in a format that suits our particular needs, helping us comprehend new information.

A strong narrative makes sure that everyone who is reading your findings can follow the intuitive path towards the insights you have found.

Writing a Report

You will now be tasked with creating a portfolio to display your work, and this document will be reviewed by your peers from this cohort.

As you develop your report, the minimum components to cover when conveying your data story are:

- Table of Content
- Project Description (Background of the project with the business problem clearly defined)
- Design (What steps have you taken to clean the data. What visualization tools have you used to share the data and why.)
- Findings (include your 3 tables from Excel and some visuals from Power BI)
- Analysis (your findings from the root cause/5 Whys analysis)
- Conclusion

Helpful Resources

5 Whys: The Ultimate Root Cause Analysis Tool (Understanding the Problem)

How to Tell a Story with your Analytics Reports

Submit Your Excel file, Power BI and your report in pdf with your name for example

“Phoebe Adjei-Frimpong - Capstone Project.pdf”

Thank you

Good Luck!!!