Strategy & Analytics Homework Assignment

Welcome to the Strategy & Analytics homework assignment! The purpose of this assignment is to provide an opportunity to showcase your analytical skills. While this assignment is representative of the subject our team's work, the accompanying data sets are randomly generated. They do not represent real customers or provide any information regarding Lucid's growth or financial state.

Problem Description

Lucid attracts visitors to its website through a variety of marketing channels, including the following:

- **video:** Video advertisements; e.g., YouTube
- **organic search:** Unpaid search engine results
- **paid search:** Paid advertisements on search engines
- **display:** Display advertisements embedded within web pages
- **partner referral:** Referrals from other applications that integrate with Lucid's products; e.g., Google Docs
- **social:** Advertisements run in social media; e.g., LinkedIn
- **email**: Email advertisements; e.g., in Gmail

These channels redirect the user to a "landing page," a web page designed to showcase Lucid's value and convince visitors to start a subscription or trial, at which point they receive access to the product.

The marketing team has been tracking site visits and subsequent subscriptions for two years now, but, up until now, no one has had a chance to thoroughly explore and understand the data. That's where you come in! We suspect that certain customer segments are more valuable than others, but we need your help to understand those differences and develop recommendations for improvement.

Data Description

You'll be working with data from two tables: visits.csv and subscriptions.csv. The recruiter should have supplied both files along with this document.
The "visits" data set records all visits to our website, including the following fields:

- **visit_id:** Unique identifier for the visit

- **account_id:** Unique identifier for the account created by this visit; note: an account is only created when a visitor starts a subscription or a trial
- **day:** Date on which the visit occurred
- **region:** Geographic region associated with the visit
- **language:** Browser language associated with the visit
- **channel:** Marketing channel responsible for generating the visit
- **landing_page:** Landing page visited; potential values include the capital letters A through J

The "subscriptions" data set contains an entry for the first subscription (if any) generated from each site visit. It contains the following fields:

- **subscription_id:** Unique identifier for this subscription
- **account_id**: Same as the **account_id** in the visits table, except this table should contain an **account_id** in every row
- **subscription_start_date:** Start date for the subscription, if applicable
- **subscription_end_date:** End date for the subscription, if applicable
- **trial_start_date:** Start date for the trial, if applicable
- **trial_end_date:** End date for the trial, if applicable
- **revenue:** Revenue resulting from the subscription in USD; note: Trials do not generate revenue, so this row is populated for subscriptions only.

Instructions & Questions

The marketing team has requested that you answer the questions in the subsections below. We expect that you will spend no more than two hours on the assignment and will grade your work accordingly, recognizing that such a limited timeframe may not be sufficient to perform the highest quality analysis.

When you are finished, please send us your answers and the file you used to conduct your analysis (e.g., your code or spreadsheet). We have created a [Google Colab notebook](#) supporting SQL, Python, & R that you are welcome to use for your analysis; however, usage of the notebook is optional. You are free to use whatever tool you are most comfortable with, but as an analytics engineer, your work will primarily be done in SQL, so we recommend you use this assignment to demonstrate your understanding of SQL. The recruiter should also have sent you the notebook in .ipynb format in case you'd prefer to run it locally.

Questions

1. How much revenue did each region generate within the time frame covered by the data sets?

2. Which channel had the highest revenue per visitor?

3. Which landing page had the highest overall conversion rate (% of visitors that go on to become paying subscribers)?

4. Did any regions have a different top converting landing page (as compared to the previous answer)?

5. Which region had the highest direct subscription rate (% of visitors that start a subscription directly without doing a trial)?

6. Which region had the highest trial start rate (% of visitors that start a trial, regardless of whether those users ultimately cancel or become paying subscribers)?

7. Which region had the highest trial conversion rate (% of trials that start a subscription)?

Rubric

While we can't divulge the full rubric, we are happy to share the general criteria and weighting for each category (see table below). Good luck!

Category	Criteria	Weight
Question 1	• Logic and answer are correct	9%
Question 2	• Logic and answer are correct	9%
Question 3	• Logic and answer are correct	9%
Question 4	• Logic and answer are correct	9%
Question 5	• Logic and answer are correct	9%
Question 6	• Logic and answer are correct	9%
Question 7	• Logic and answer are correct	9%
Organization & code quality	• Code is well written and easy to follow • Code approach is reasonably performant	37%