You have been given two files:

1. Case Task: The document that you are reading now, which includes data definitions
2. CASE_STUDY_DATA: A table of performance associated with editorial content published on the VICE website in a recent period

Please ensure that you read this entire document before beginning the exercises.

Provide your results in a new document(s). Please ensure to label each answer with which question it pertains to.

Part 1

In this section of the case study, we ask you to write several queries to evaluate your SQL skills and ability to work with a new dataset. We've done our best to make these questions as relevant as possible to real-world tasks that will be done on the job.

We have provided enough information here to successfully complete all of the tasks, but we understand questions may arise. It is okay to make reasonable assumptions about the data so long as you provide documentation of the decisions being made. Please pull the provided dataset into your preferred tool for analysis and then complete the questions below.

Each unique article is represented by a six-character, alphanumeric ID in the URL string. For example, 4avnan is the article_id here:
www.vice.com/en_us/article/**4avnan**/bombshell-report-finds-phone-network-encryption-was-deliberately-weakened

1. Use an SQL query to extract the article_id from the URL. Please paste the query in the box below. You may also link to a separate sheet/document if necessary.
2. Using the published_date field, calculate the unique daily and weekly article output. How does content output trend throughout the week?
3. VICE's editorial department is structured into different sections (e.g. Motherboard, US Culture and Horoscope). How does each section contribute to editorial output?
4. Write a query creating a table that shows:
    a. each section
    b. its unique content output
    c. its traffic contribution
    d. Section's share of the total output and traffic over the full data period.

Part 2

In this section of the case study, we are looking to assess your skills in deriving and communicating insights from data. We are also asking you to provide some qualitative commentary to demonstrate how you would think about content performance. The purpose here is not to find a single correct answer, but rather to show how you approach thinking about these questions.

You can either continue working in SQL or transition to Excel or Google Sheets. In this role, we are looking for a candidate who has a fundamental understanding of SQL and is able to Google their way through difficult questions.

In this dataset, content traffic comes from four traffic_channels: *Social*, *Search*, *News Aggregators*, and *Other*. The traffic source mix varies both at the individual article and section level.

1. Create a chart(s) showing the traffic channel distribution split by section. What do you think is the value of each traffic source, and how would you use them to assess performance?

You've probably noticed that the dataset occasionally contains multiple URLs for each article. These represent the different language publications of the same article. The locale field indicates the language or country with which an article is associated. For example, the locale fr represents French language content and the locale en_us represents English content from the US.

2. Using your preferred mix of tables and charts, give a sense of how content output and traffic is distributed across the locales and languages. Remember that despite sharing different locales, en_us and en_uk belong to the same language.
    a. Is there a least / most effective language for content to be published in? Is there a difference in traffic source mix depending on language? Does the number of translations for an article appear to be related to its total traffic?

3. For the last part, we want to issue a recommendation on future content investment. Using the sample data provided, group the content by section and provide quantitative *and* qualitative recommendations about how VICE should increase output most effectively from a traffic standpoint.
    a. This is an open-ended question that can build on the insights obtained in the prior steps. While the data as given is split by section, you should feel free to further segment the content based on language or topic. For example, there are a lot of articles within the US Culture section that contain recipes, and these can be segmented out.
    b. In this part, you do not need to create a data framework explaining all 1,183 articles. However, if you are making the case that a topic, such as recipes, should adjust output up or down, then try to include a metric to support your point. In this simplified scenario, the goal is to maximize pageviews and minimize inefficient content output.
    c. Don't spend too much time on this last part, the goal here is to get a sense of how you approach thinking about and justifying content recommendations.

That's it! This case study should not take more than 4 hours to complete. If you have any questions, reach out to laura.lloyd@vice.com. When you are comfortable with your results, submit to your recruiting partner at VICE Media.

We're looking forward to seeing your results!

Data Definitions

- original_article_title: The original, English language title of an article
- url: The URL of an article
- locale: The URL fragment indicating the language and/or country to which an article was published
- article_id: The unique six-character identifier for each article
- published_date: The date an article was published
- section: The editorial business unit that created this article originally
- traffic_channel: The traffic source from which a user arrived at an article
- page_views: The number of page views to an article