Data Science Case Study

Overview:
Thank you for taking the time to interview for a Data Scientist role at Apple! This is a take-home case study created to assess the candidate's core Data Science skills and strategic mindset. The candidate should have at least 7 calendar days to prepare for the case study and the subsequent presentation.

Case Background:
Pippin is a San Francisco-based startup founded in 2017. The firm has been a disruptor in the music streaming industry. The Pippin app allows users to listen to hundreds of million songs at an affordable monthly price. You work as a data scientist at Pippin. In the last few weeks, Pippin has launched an A/B test to optimize subscriber conversion. You have been asked to analyze the test results and to deliver a short presentation to the broader project team including the members of Product Management, Marketing, Finance, and Engineering. Before meeting the broader team, you need to review your work with your manager and peer data scientists on the team.

Assignment:
- Analyze the accompanying dataset (*Pippin_subsciptions_data.csv*) and develop actionable insights
- Create a 15 mins presentation that includes
    - Overview of subscription conversion trends
    - A/B test results
    - Recommendations for the A/B test and overall subscription growth
- Be prepared to answer questions that your manager or peers may have during the review session. Your team is comfortable with looking at code if needed. However, the broader project team prefers a business presentation.

Deliverables:
Please email the following documents at least 1 day before your interview to your recruiter:
- A presentation deck in pdf format.
- Code or data process document in csv/notebook/workbook format.

Interview Agenda:
- The interview will be set up as a review session with your manager and peers
    - **15 mins**: Present the deck you prepared for the broader project team
    - **15 mins**: Walk through the additional details of your analytic approach, assumptions, code, process, etc.
    - **30 mins**: Q&A and discussion

- Please expect a collaborative discussion during the interview. You may want to have your analysis environment ready to answer the questions that may come up during the discussion.
- For virtual interviews, we will use Cisco Webex for video conferencing and screen sharing.

Data Dictionary (Pippin_subsciptions_data.csv):

- customer_id: each row represents a unique customer
- click_date: the date on which the customer clicked on the Pippin banner to express interest in the subscription service
- platform: platform used by the customer (desktop web or mobile web)
- customer_city: city from customer address
- customer_state: state from customer address (US only)
- urban_flag: 1 → customer lives in an urban area, 0 → otherwise
- credit_card_on_file: 1 → customer has a credit card on file, 0 → otherwise
- student_flag: 1 → customer is a student, 0 → otherwise
- fav_genre: customer's favorite music genre obtained from 3rd party data-source
- subscriber: TRUE → customer enrolled for Pippin monthly subscription, FALSE → no subscription
- variant_id: 1 → control group, 0 → test group

Additional Information:

- This is a hypothetical dataset created for the sole purpose of data science interviews. It does not reflect any data collection policies at Apple.
- Assume that all the available features related to the A/B test are included in the dataset.
- The presentation with the broader project team is scheduled for 15 minutes but there is no limit on the number of slides in the presentation.
- There are multiple approaches you can take to solve the problem and there is no single right or wrong answer. Please use your best judgement for selecting the necessary methodologies and recommendations.
- Your main immediate task is to present your analysis to your team and get their buy-in on the presentation to the broader team.
- The case study represents 1-day work for a data scientist in our team. You can take as much time as you would like to complete the case study.
- Feel free to use any tools, languages, or programs. Most data scientists use a combination of Python and SQL. At Apple, we have robust internal support for R, Scala, Julia, Java, Tableau, and more.
- For the sake of simplicity, we have given you a small sample i.e. ~10k rows and a handful of columns.
- Please do not consult or share this case study with any other person.

Contact Information:

If you have any questions about this exercise, please contact Erick Cai (ecai@apple.com) or Vishal Morde (vmorde@apple.com). Good luck!