

Background:

An internet service provider is struggling to optimize its product portfolio. Over time, different product managers have created various products to address competitive offers, experiment with customer responses, and more. This has resulted in a challenging situation for customer service agents who must remember numerous products, for technicians who must track various product configurations, and even for the product managers themselves who must design public communications and promotional materials. The number of personnel required to manage these operations, the platform and system resources needed, and the promotional budget are all increasing disproportionately to the marginal growth in revenue. The Head of Finance has expressed concern about this issue to the Head of Products, who agreed to review the product portfolio.

The Head of Products tasked the product managers with reducing the number of products in the market without harming revenue. After a week of effort, they proposed a streamlined product portfolio. However, the Head of Products is uncertain about accepting the proposal due to concerns about its revenue impact, despite assurances from the team based on their extensive marketing experience.

As the Chief Data Scientist, Head of Products calls you over phone on Friday 9AM. He briefs you the above and asks you to provide your perspective on whether to proceed with the proposal. Your feedback will be crucial as the Head of Products needs to present this proposal to the CMO, CFO, and CEO for approval at 5PM Saturday.

After consulting with your Data Science team, you determine that the following questions must be answered to assist the Head of Products in making an informed decision on the new product portfolio:

- A. What will be the likely uptake or purchase rate of the new products in the portfolio?
- B. Which customers are most likely to convert to the new products for targeted campaigns?
- C. In which locations should new product promotions be run?
- D. Is there any potential for optimizing the proposed product portfolio further?

You need to gather all relevant information, supported by data, summarize your insights, and present your recommendations to the Head of Products convincingly.

Keep in mind, CXO presentation is at 5PM. Prior to that, you need to align Head of Products on your recommendation by 2PM Saturday so that he can prepare on time for his presentation. You, yourself, need to review the findings of your team by 12PM Saturday so that you can summarize for Head of Products alignment on time.

Data that you have in your respective S3 bucket:

a. TBL_CUSTOMER_PROFILE.csv

This dataset contains information regarding customers of the service provider which has below information.

Column Name	Description
CUSTOMER_ID	Customer ID
SERVICE_TYPE	Standard/Premium Customer
FIRST_CONNECTION_DATE	From when using this connection
CONNECTION_TYPE	What types of connection customer use
ROUTER_CATEGORY	Customer's Router category
STREAMING_VOL_PERCENTAGE	Streaming Volume as % of Total data consumption
WEB_BROWSING_VOL_PERCENTAGE	Web Browsing Volume as % of Total data consumption
IM_VOICE_CALL_VOL_PERCENTAGE	IM or Voice Call Volume as % of Total data consumption
SNS_VOL_PERCENTAGE	SNS Volume as % of Total data consumption
FILE_ACCESS_VOL_PERCENTAGE	File download Volume as % of Total data consumption
GAME_VOL_PERCENTAGE	Online Game Volume as % of Total data consumption
NAVIGATION_VOL_PERCENTAGE	Navigational App/Web Volume as % of Total data consumption
EMAIL_VOL_PERCENTAGE	Email Volume as % of Total data consumption
OTHERS_VOL_PERCENTAGE	Other usage Volume as % of Total data consumption

b. TBL_PRODUCT_CATALOGUE.csv

This table contains product packages available for customer to purchase.

Column Name	Description
PACK_ID	ID of the pack
SERVICE_TYPE	Standard/Premium subscriber's pack
PAYMENT_METHOD	Payment method used to purchase this pack
PACK_PRICE	Pack Price
VALIDITY	Pack Validity
DATA_VOL_GB	Data volume (in GB) for this pack
PRODUCT_STATUS	Discontinued/Continued/New

c. TBL_PACK_PURCHASE.csv

This table contains history of pack purchase for each customer.

Column Name	Description
EVENT_DATE	Pack purchase date
CUSTOMER_ID	Customer ID

PACK_ID	Pack ID
HITS	Number of purchases

d. TBL_DATA_CONSUMPTION.csv

This table contains customer daily data consumption information.

Column Name	Description
EVENT_DATE	Date of Volume Consumption
CUSTOMER_ID	Customer ID
TOTAL_VOLUME_GB	Total data volume (GB) consumed
VOLUME_GB_IN_HIGH_BANDWIDTH	Data volume (GB) consumed with high bandwidth

Questions to answer, submission guidelines, assessment criteria:

Q1: You need to prepare a machine learning model to predict out of the given customer list in 'Q1-customers.csv' who will purchase which of the new products. If any customer is not likely to purchase any of those put -1.

Sample submission file is given as 'Q1-submission.csv' in your respective S3 bucket. You need to update the file with your prediction. We will read this file periodically remotely and update leaderboard based on accuracy from 4PM Friday to 4AM Saturday. First submission has to be by 4PM Friday and Final submission time for this question is 9AM Saturday. Late submissions will be disqualified.

Your submission will be assessed based on i) Accuracy, ii) Feature Engineering iii) Documentation iv) Coding Standards v) Technical sophistication vi) Explain ability

Send your final submission as teamcode-Q1.csv and code files as teamcode-Q1-*.py or teamcode-Q1-*.ipynb or relevant ././rmd files to datathon@robi.com.bd.

Q2: You need to predict probable uptake or purchase count of the products given in 'Q2-new-products.csv'.

Sample submission file is given as 'Q2-submission.csv' in your respective S3 bucket. You need to update the file with your prediction. We will read this file periodically remotely and update leaderboard based on RMSE from 9PM Friday to 9AM Saturday. First submission has to be by 9PM Friday and Final submission time for this question is 12PM Saturday. Late submissions will be disqualified.

Your submission will be assessed based on i) Accuracy, ii) Feature Engineering iii) Documentation iv) Coding Standards v) Technical sophistication vi) Explain ability

Send your final submission as teamcode-Q2.csv and code files as teamcode-Q2-*.py or teamcode-Q2-*.ipynb or relevant .r/.rmd files to datathon@robi.com.bd.

Q3: You need to propose top 10 places where to place billboards for promoting these new products across Bangladesh. Use any public data source to support your recommendation.

First submission has to be by 6AM Saturday in your respective S3 bucket as teamcode-Q3-*.jpg or teamcode-Q3-*.gif and Final submission time for this question is 12PM Saturday. Late submissions will be disqualified.

Send your final submission i) image file of the Bangladesh map as teamcode-Q3-*.jpg or teamcode-Q3-*.gif with marker for those places ii) source code to extract those data from public source as teamcode-Q3-*.py or teamcode-Q3-*.ipynb or relevant .r/.rmd files to datathon@robi.com.bd.

Your submission will be assessed based on i) data source used – its granularity, cost and reliability ii) Any other utility of your shared data source – please articulate those utilities in your source code file.

Q4: You need to share your recommendation on what an ideal product catalog should be for “Premium” customer with “Cash” payment and for 30 days validity. Submit 10 products that needs to be added as ‘Q4A-submission.csv’ (sample file given) to the product list and 10 products that can be dropped as ‘Q4B-submission.csv’ (sample file given) from new product.

First submission has to be by 12PM Saturday and Final submission time for this question is 4PM Saturday. Late submissions will be disqualified.

Send your submission as teamcode-Q4A.csv/ teamcode-Q4B.csv and code files as teamcode-Q4-*.py or teamcode-Q4-*.ipynb or relevant .r/.rmd files files to datathon@robi.com.bd.

Your submission will be assessed based on i) Methodology and ii) Accuracy. Add necessary documentation for assessors’ understanding.

Assessments:

- i) Walk-in-assessment: At random intervals organizing team will monitor and interact to understand your progress, methodology, business and tech understanding which will be part of overall assessment.
- ii) Jury board: Top 15 teams will be selected based on final submissions till 12PM Saturday. You will need to present your works based on your submission till 12PM Saturday to jury board. They will assess you on i) Technical sophistication ii) business understanding and iii) communication.

Number of Board: 3, Team per board: 5

Time limit: 10-minute (sharp) presentation + 10 minute Q&A.

Venue: L19 and L15 meeting room

You need to submit your presentation material as Teamcode-*.ppt by 1PM Saturday at datathon@robi.com.bd.

- iii) Grand jury: Top 7 teams will be selected based on Jury board assessment and they will present all their submission up to 4PM Saturday to Grand Jury. They will assess you on i) Business understanding and ii) Presentation.

Time limit: 5-minute (sharp) presentation + 5 minute Q&A.

Venue: L11

You need to submit your presentation material as Teamcode-*.ppt by 4PM Saturday at datathon@robi.com.bd.