



SENIOR DATA ANALYST - BUSINESS CASE

We're excited to have you join us on this journey! To better understand the role and its responsibilities, we've prepared two case studies for you. These will provide insight into the type of work and challenges you'll be facing as a Senior Data Analyst at Xapo Bank. Let's get started!

Case Study 1: Expanding a Crypto Bank Offering a Global Card

Background

Cryptosphere Bank offers a global debit card that allows customers to spend cryptocurrencies and stablecoins (BTC, ETH, USDC and USDT) anywhere in the world. The bank was launched a few years ago and is now looking to accelerate growth and expand its operations.

Objective

As a Data Analyst, your task is to analyze customer data for patterns and insights to identify valuable customer segments and to also identify potential new markets for Cryptosphere Bank. Please incorporate the following requirements into your analysis.

- Perform an exploratory data analysis to identify patterns, trends and insights that can help with understanding the customers behaviour and segments, the card performance across regions and also with identifying potential new markets. Consider using advanced analytics techniques for identifying customer segments. Aim for the exploratory data analysis to be self-explanatory.
- Conduct hypothesis testing to identify if there are significant differences between customer segments, regions, or countries.
- Based on the exploratory data analysis findings, develop an interactive dashboard for stakeholders to provide an overview of customer spending behavior and potential new markets for the bank to expand its operations.
- Design an A/B test(s) to validate the impact of the change(s) you're recommending.
- Ensure that all code and data are clean and easy to understand.

Datasets

You are provided several datasets containing the customer data and related dimensions. Feel free to use data from external sources in your analysis. If you do so, please be explicit about it and explain the reasoning behind using them.

The provided datasets are explained below.

customers

This dataset stores information about Cryptosphere Bank customers.

- `customer_id`: a unique identifier for each customer of the bank.
- `creation_date`: the date and time when the customer's account was created.
- `first_approval_date`: the date and time when the customer's account was first approved by the bank.
- `residence_country`: the country code (alpha-2) where the customer currently resides.
- `passport_country`: the country code (alpha-2) where the customer's passport was issued.
- `date_of_birth`: the customer's date of birth.

- gender: the customer's gender.
- occupation: the customer's occupation.
- annual_income: the annual gross income in USD of the customer.

[cards](#)

This dataset contains information about Cryptosphere Bank's credit and debit cards.

- card_id: a unique identifier for each card issued for the bank customers.
- customer_id: the unique identifier of the customer associated with the card.
- type: the type of card (e.g. credit, debit).
- subtype: additional information about the card type (e.g. plastic, virtual, metal).
- status: the current status of the card; the accepted values are ready_to_activate, request_cancelled, active, closed.
- brand: the payment network the card belongs to (e.g. Visa, Mastercard).
- currency: the currency of the account the card is attached to; the possible options are BTC, ETH, USDC and USDT.
- creation_date: the date and time when the card was requested.
- activation_date: the date and time when the card was activated.
- closure_date: the date and time when the card was closed (if applicable).
- paused: it indicates whether the card is currently paused or not.

[card transactions](#)

This dataset contains information about transactions made using Cryptosphere Bank's credit and debit cards.

- transaction_id: a unique identifier for each card transaction.
- card_id: the unique identifier of the card used for the transaction.
- created_at: the date and time when the transaction was created.
- transaction_type: the type of transaction; the accepted values are generic_card (debit transactions), generic_reversal (credits to cardholder's account associated with a previous transaction), cash_withdrawal (cash withdrawal from ATM), chargeback (transaction resulted from a chargeback procedure)
- status: the current status of the transaction; the accepted values are pending, completed, rejected, cancelled
- payment_method: the payment method used for the transaction; accepted values: contactless, e_commerce, card_on_file
- mcc: the Merchant Category Code for the merchant where the transaction took place.
- amount_usd: the amount of the transaction expressed in USD.
- currency: the currency of the transaction.

[mcc-codes](#)

List of Merchant Code Category (mcc) codes along with their descriptions

- mcc: the numerical code associated with the Merchant Category Category (MCC) for a particular business or industry. MCC codes are used by payment processing companies and financial institutions to identify the type of business that is processing a payment.
- description: brief description of the industry or type of business that is associated with a particular MCC code.

[country-codes](#)

This BigQuery public dataset contains the codes for the representation of names of countries following the ISO 3166-1 standard.

- country_names: the names of countries in English.
- alpha-2: the two-letter country codes defined by ISO 3166-1.
- alpha-3 encodings: the three-letter country codes defined by ISO 3166-1.

Deliverables

- a brief document explaining your approach, choices, thought process, interpretation of results and insights gained from the analysis. Please include recommendations and clearly list any assumptions you made and the limitations of the chosen methods, or techniques.
- the exploratory data analysis.
- the dashboard.

Case Study 2: ETH Wallets

Background

Ethereum is a crypto currency which leverages blockchain technology to store transactions in a distributed ledger. A blockchain is an ever-growing "tree" of blocks, where each block contains a number of transactions.

Objective

Use the [Ethereum Bigquery public dataset](#) to answer the following questions using SQL:

- Which are the top 10% addresses in terms of transaction volume in the past 6 months? What can you tell about their transactional behavior? Classify them in spenders and accumulators.
- Which are the top 5% wallets by balance? For each of them determine:
 - first transaction date
 - last transaction date
 - monthly transacted volume in ETH
 - transaction frequency
 - total number of transactions
 - total transacted volume in ETH

Deliverables

- the answers to the questions
- the queries used.

To share the deliverables for the two case studies, consider the following options:

- private Github repo shared with the hiring manager (her Github username is andreeaxapo)
- Github Codespaces
- archive with all the files send via email to Gabriel (gabriel.fistos@xapo.com)

We're looking forward to learning about your insights and recommendations!