

Welcome to the Extend analytics exercise, we appreciate your time and effort in completing this. This exercise will help Extend focus on candidates who are best suited for the analytics role, and also helps the candidate get an idea of the type of data we work with and the technical skills required to be successful in the role.

Setup

This exercise is designed to gauge your skills in working with data, particularly in 3 key areas:

1. Data cleanliness and integrity
2. SQL design, factoring, and querying
3. Data analysis and visualization

You are free to use any tool/programming language you're comfortable with. The methods utilized for this exercise are meant to emulate the kind of work we are doing on the Risk & Analytics team, and therefore SQL and Python/R skills will be valuable for being an effective contributor. From the provided CSV files, you will import the data into Jupyter/RStudio, correct any data integrity issues during the process, and utilize SQL to design a schema and tables for querying the data.

If you are skilled in both languages or do not have any software preference, we recommend Jupyter notebook and SQLite. Jupyter is typically run from within an Anaconda installation, and this can be downloaded from <https://www.anaconda.com/products/individual> if you do not have it already. Alternatively, Jupyter can be installed by command line with Python or Miniconda installations, and instructions can be found at <https://jupyter.org/install>. If you need help creating a local SQL instance to run queries, please follow these [instructions](#) for SQLite.

When preparing the deck to answer the questions, making it fairly readable as a standalone item is best but also be prepared to walk us through it.

If you need assistance in getting Jupyter up and running or have any questions about the exercise, please reach out to Alex Raya at alex.raya@extend.com.

Data Overview

- orders.csv
 - Purchase data for Extend's Shopify merchants.
 - Check it out at a merchant like [August](#) to get a sense of the data.
- order_lines.csv
 - Individual line items purchased within an order.
- contracts.csv
 - Each warranty contract purchased by a consumer.

- merchants.csv
 - Data on each merchant that has integrated Extend's purchase protection product.
- products.csv
 - List of all products sold by merchants, with product details and whether or not they are warrantable.

Questions

1. Please review the integrity of the data. Do you notice any data anomalies or idiosyncrasies? If so, please describe them.
2. Calculate attach rate overall and by merchant per month on a unit and dollar basis. Pick two merchants that are outliers (high attach rate and low attach rate) to do a SKU level analysis.
3. Which merchant industry would you focus business development on based on current attach rate (by product category or price point) and your views on market sizing? Assume we have roughly the same market penetration in each industry so that saturation is not a concern and assume net revenue to Extend is calculated by warranty sales * (1-merchant cut). Please put together a 3-page PowerPoint presentation to the Extend executive team with your recommendation (title and agenda slides do not count in the total).
4. After exploring the data, define at least two questions you would be interested in exploring. Include 1 additional slide at the end of your PowerPoint presentation as "Next Steps" and at a minimum, supply:
 - a. What the question is and why does it matter for the business.
 - b. What in the current data you have assembled makes you interested in the question.
 - c. What additional data would you want to explore the question.
 - d. At a high level, what would you do with the data to answer the question.
5. Please send a Jupyter/Rstudio/SQLite notebook with your work and the files you utilized for completing the exercise.
6. Please share any feedback you have on the exercise and how we can make it better!

Grading/Evaluation

The Risk & Analytics team will be evaluating the completed assignment across the following dimensions for Q1-Q4

- Answers / Presentation
 - Is it typo free?

- Are slides organized in a way that's easy for reviewers to follow?
 - Are questions properly answered?
- Explanations
 - Are explanations clear and easy to follow?
 - Are explanations complete but concise?
- Coding
 - Is the code structured properly?
 - Is it written in an efficient way?
- Documentation
 - Is there sufficient documentation?
 - If you were to ship this off to another team member, would he/she understand what the code does?