Excel Inc Mobile Data Case Prompt

In this case, you work for Excel Inc., a (fictitious) home improvement store. You have been assigned to work on an analytics project involving mobile data licensed from a data vendor. The primary goal of the project is to measure shopping behaviors.

Excel Inc has recently started working with a vendor that provides them with location data on shoppers of home improvement stores. This data includes all individuals who are measured as having visited a home improvement store and loosely indicates how often, where, and how long a consumer is measured as being at the home improvement store. The data covers visits to home improvement stores of Excel Inc and their key competitors. The full data covers approximately a year of visits by a large sample of tens of millions of consumers.

Broadly, this project is intended to start a new process of measuring shopping behaviors of consumers. The intent is to understand how many shoppers each home improvement store chain gets, how loyal consumers are in terms of shopping, and how much cross-shopping is happening. The project sponsor will provide more detail.

You will make a presentation of this project that focuses on the current state of these shopping behavioral measures. Ultimately, the idea is not to do a single report, but to create a data product that will be updated regularly and allow on-going reporting and analysis using the measures developed in this project. To serve this goal, you need to design a data warehouse that contains the business-relevant information about the cross-shopping behavior.

The data warehouse is intended to serve a wide range of analytics needs. Again, the sponsor will provide more information, but these needs include assessing the effect of a variety of marketing actions (loyalty programs, promotions, advertising, pricing) by both Excel Inc and competitors. As a result, the goal is to design some measures that capture important shopping behaviors and segments of consumers. Users of the data will need to be able to understand the measures, so simple ways of describing them are needed, even if you use sophisticated methods to construct the measures.

The project sponsor will give more details about the mandate of this project. However, broadly, your task in this project has three components:

1) Design the analytics project and propose the design to the sponsor as a brief set of "dummy" analysis slides (no recorded presentation is needed)
2) Create of a data warehouse (multiple options for how to complete this stage)
3) Analyze the data and present the first report about consumer shopping behaviors

Audience

Excel Inc has a number of relatively sophisticated analysts as well as quantitatively trained MBAs, a centralized pricing group, and a number of other reasonably sophisticated individuals. The project sponsor, Han Ellickson, leans into analytics, but is not actually trained deeply on analytics, having grown up in a spreadsheet world. He understands the big picture, which is why he thinks this data could be important. In fact, despite some relatively new staff that are quite strong in analytics, the bulk of managers that will interact with this report and use this data do not have much sophistication in analytics. Due to this variation in sophistication, the presentation and data dictionary should be accessible to this broader audience.

That said, the up-front design of the analytics doesn't need to be "dumbed down" so much as made reasonably clear. The data warehouse and underlying data work, in particular, does not need to be accessible to the broader audience. In fact, the more the data warehouse can hide the sophistication, the better.

Kick-off meeting

The project sponsor, Han Ellickson or one his associates will meet with you to kick-off this project. You can write emails to Han at han.ellickson2u@gmail.com. You should look for an email from Han to kick off the project.

The email will contain important information about the project mandate. In addition, you will have a video to watch that represents the first part of the kick-off meeting with Han or one of his associates. The live (over Zoom) part of the meeting is your chance to ask clarifying questions of the project sponsor.

To be clear, you should clarify in your mind what is a question to the instructor/TAs (e.g., clarifying the parameters of the assignment such as deadlines, optional vs. requirements for the grading, etc.) vs. questions to the sponsor. The instructor questions you can ask in class, during office hours, or by email to the instructor.

Questions to the sponsor can happen by email or during the kick-off meeting. The project sponsor will be available for emails after the kick-off meetings, and you can ask for a follow-up (short) live meeting. He or his staff may or may not be available for such meetings given their busy schedules and will do their best to respond to your emails.

Stage 1: The Design Agreement

Given the scope of this work and the involvement of a costly data vendor, the first step in this project is to get clear about the design agreement. This involves submitting a proposal of the planned measures and analytics as a brief "dummy" analysis and data design presentation.

The proposal presentation should spell out the following:

1) The data design for the data warehouse (all entities and their respective attributes) and planned idea of how you will construct the data warehouse. This should be illustrated with a slide or two. Typical slides here would be a visual of the tables with clearly named attributes and a visualization of the approach to the data construction (i.e., intermediate file, direct access, or exporting).
2) Analytics outputs or simple analysis plans related to constructing the report from the data. This can involve example graphs or tables with "dummy data" much like those in the Astro Beer case. For more sophisticated analyses, you can briefly/concisely describe the plan.
3) Importantly, the proposal should be clear about any data definitions such as new measures.

Each of these design elements are discussed in more detail below for the final assignment. You will need to interact with the project sponsor to obtain all of the refined information for the design. Submit the presentation as the proposed project agreement and then the sponsor will respond with a final, approved agreement. This approved agreement may differ meaningfully from your proposal, so be sure to follow it carefully.

The data vendor will provide the dataset as well as some information as background. Although you will have access to the data, I don't recommend digging too far into the analysis until you receive the agreement on the plan for the project. You should engage with the data to get familiar with it and to help you get concrete about the analysis and data designs. The data will have both an .Rdata file version and a MySQL version (available later).

Your final analysis and presentation should aim to leverage the data warehouse you build based on the full dataset.

Stage 2: The Data Product, Analysis, and Presentation

After the sponsor sends the approved project plan (which might differ from your proposal!), you will need to work with the data in one of the two formats provided. The first is an "intermediate file" approach where you are given the .Rdata files containing all of the data. By using these files for your final data work, you are not creating the actual data warehouse, but working from a data dump. The second is a direct access approach, where you can access a MySQL database that contains the data. Either can work to achieve the assignment, but extra credit will be awarded for solutions that programmatically connect to the MySQL database.

Once you have the approved project plan, you should proceed to write and refine your code for the data warehouse and analysis. You will ultimately make a presentation about the shopping behaviors. The data warehousing and wrangling aspects will be graded separately by Liza Mohr in the assignment you submit to her.

The Data Product: A Shopping Behaviors Data Warehouse

For the data warehouse design, you need to identify the level of aggregation (i.e., what constitutes an entity) as well as the attributes of that entity (the columns in the data table). The expectation is that you will want more than one table/entity. You will need multiple tables to address the variety of requests stakeholders might have as well as your own report.

More broadly, the design of the data warehouse should consider efficiency, simplicity of use, and storage requirements. Do not create an overly complex database with many, many tables. A few simple tables that can capture the key uses of the data is what is expected. Hence, you should balance the various factors to create your data warehousing product.

For the graded assignment, you have three options of how to construct the data warehouse.

- **Option A (Sophisticated Option)**: Use only the .Rdata files of the data. Use R statements to construct the interim data (e.g., as data.table). This would involve creating a .Rmd file that has the R statements clearly commented.
- **Option B (More Sophisticated Option)**: Use the MySQL database and access-only queries that pull data from the MySQL database (not all of it!). Do this via SELECT statements and the interface in MySQL Workbench that allows exporting/downloading data. Then use this downloaded data to construct your data warehouse. Provide a document with the SQL queries to extract the data in the format for data warehouse.
- **Option C (Most Sophisticated Option)**: Use federated data approaches to create the data warehouse by directly linking to the MySQL database and constructing a local MySQL database as the warehouse. Provide a document that contains the code used to construct the data warehouse.

You choose one of the above options to submit. In addition to the above code approaches, you will also need to provide a data dictionary that describes each data set you develop and plan to store in your data warehouse (data.frame, data.table, or table in a mysql database) and attribute in the table (i.e., column).

Analysis and Presentation: First Report on Shopping Behaviors

The analysis and presentation are intended to be the first of a regular reporting process on shopping behaviors. Further, as Excel Inc and competitors take actions, the firm will track how these actions impact shopping behaviors using this new data source. This is in some sense similar to other tracking efforts through surveys (e.g., brand tracking, customer satisfaction), social media listening, campaign analytics, website analytics, and search analytics, but less involved at this stage.

The analysis should center on the key business needs identified through the agreement with the project sponsor. Likewise, the presentation should serve to meet the business needs identified in that document.

However, as the first presentation using this new data, the data and measurement should be highlighted. This involves being clear about the key measures that have been developed and how/why they are useful for monitoring shopping behaviors. Perhaps one or two slides could be used for this purpose. Such measures can include dimensions related to creating groups of customers as well as ones that describe shopping behaviors for a given group of customers. For this purpose, you will likely want to use methodology slides that describe the spirit methodology as simply as possible for non-technical people. Ideally this is done visually.

The presentation should aim to be 5 minutes along with Q&A afterward. The presentation doesn't have an explicit format, but should match the time limit for the presentation. You should leverage highlights and detail slides effectively in your presentation.

In addition to the presentation, you should submit any relevant analysis code as a .Rmd file. This can include multiple parts. First, we want the presentation to be reproducible, so that the code should either create the data used to make the charts (if in Excel or other tool) or the charts themselves (if created in R). Second, any code to develop measures should be clearly indicated. The code should be commented as to be sufficiently clear.

You will also receive a more detailed specification from Liza Mohr for the structure for the data deliverable for her class.