GovEx

Quantitative Analyst
Interview Exercises

Instructions

Please take no longer than three days (72h) to complete the following exercises and return the answers, scripts, or notebooks used to explore the data, and the presentation slide deck to mvaugha6@jhu.edu. In general, these exercises are not specifically designed to have a right or wrong answer, but more importantly to allow you to explore and be creative. We are more interested in the process rather than the final answer.

Part 1: Mapping questions to data

Recommended time: 1h

The City of Paterson wants to understand what is limiting business growth in the City. A recurring question that city leadership often asks is: how can the city be transformed into a thriving, liveable city? Data compilation and analysis is needed for the City of Paterson to evaluate efforts in revitalizing the city's main streets and improving economic activity.

During an initial meeting with the Economic Development Office they created a list of questions:

- Are application permits being processed efficiently?
- What is the trend in opening/closures of businesses?
- Are existing businesses faring well?
- Do residents feel that the City is safe and liveable?
- Are Public Safety response times decreasing?
- Are shootings (fatal & non-fatal) decreasing?
- What is the trend for other crime types?
- What are the locations?
- Are vacant, abandoned and foreclosed properties dealt with in a timely manner?

In order to explore the current landscape, create hypotheses of the causes that might be driving business and people outside of the downtown area, and create a performance program to track the evolution of the issue, the City has shared with us a series of datasets:

Dataset name	Description	Metrics
Vacant-Foreclosed-Abandoned Properties	List of vacant, abandoned and foreclosed properties	Property type (vacant/foreclosed/abandoned), address, violation date (2014 to 2020)
QoL_calls	Summary of Quality of Life calls received by Public Safety for each month of 2020.	Call type (barking animal, blocked driveway, noise complaint, etc), number of calls, month.
Annual report Board of Adjustments	List of residential and commercial units that come up in front of Paterson Board of Approvals	Date of the meeting (2017-2019), applicant name, address, phone number, block number, lot number, number of units, commercial type
Incidents Non-Fatal shootings	List of non-fatal shootings during 2020	Date, address, hour
911 Calls for Service	911 log of calls	Date, call hour, call type, status, closed date, last activity date, outcome, latitude, longitude, zip code, address.

Our first step will be to quickly inspect the data in order to provide an assessment of what can and can't be done as well as following steps. To do so, please answer the following questions:

1. For each dataset, what metrics would be useful for our problem? Map[1] each question to the metric that might be used to answer it. Describe in detail what the analysis should be, as well as any assumptions you made or further questions you have about the data.

2. What other metrics would ideally help the City of Paterson track the evolution of their efforts to revitalize the downtown area? Create a new list of metrics and datasets we can request the city to collect and/or share.

If this is your first time working with any government data, you can check the Baltimore Open Data portal (https://data.baltimorecity.gov/) as an example of datasets that a city typically collects.

[1] Mapping research questions to potential data points is one crucial pre-analysis step. By "map," we mean outline / describe plans for analysis by discussing potential data and data points that can be used to answer each research question.

Part 2: Data exploration and manipulation

Recommended time: 4h

The City of Syracuse approached us to build a predictive model of the house market sale prices for all the properties in the city based on the physical characteristics of the building and other metrics. They shared with us four different datasets (found here) :

- Sales
- Improvements
- Residential Building
- Assessments

The City also provided data dictionaries and some tips about the data:

- A parcel can contain several sites. Sometimes each site is a property, but other times it's only an addition that will be sold together with the main house.
- Once an improvement is built (either at the same time of the construction or later), the improvement is registered each time that there is an inspection of the property.
- Improvements are registered independently of the time of a sale. For the same property, we might find improvements done both before and after a sale, but we only want to take into account those that affect the sale price.
- There is a field in the sales table, ARMS_LENGHT (spelling corresponds with the original city data), that provides information on whether or not the sale represents an arms-length transaction. Therefore, it can be used to filter the sales considered not representative.
- There are also sales that, despite being marked as arms_lenght transaction, might not be. For example: Suspicious sales happen when a house is bought, refurbished, and sold again within a small time frame, or when the sale price is very low.

We want to explore the data, and properly combine the datasets so that we have all of the information in a unique table. The final table should have a unique sale in each row, with all the building characteristics and improvements remarkable to that sale in the same row, and should be ready for modeling (all values numeric). To prepare a final table:

1. Explore the data on each of the Sales, Improvements, and Residential Building tables.
2. Clean the data on the previous tables: look for duplications, missing or incorrect data, data type issues, necessary filters, etc. If you had to make any assumption during that process, it’s ok! Take notes and share them with us.
3. Combine the previous three datasets in a single table:
    - Merge the datasets.
    - Do any final cleaning with all the information combined, if necessary.

- Make sure each row in the final table is a unique sale (we want to predict sale prices).

We also want to explore the assessments done by city inspectors and check whether there might be any equity issues with the assessment process. Combining the assessment information with the sale prices of properties, create a graph with sale prices on the x-axis and the difference between assessment value (full market price) and sale price on the y-axis. Briefly comment what you see with an equity angle.

Part 3: Presentation

Recommended time: 1h

Imagine you are an analyst within GovEx who is going to present the results of part 1 and part 2 to the heads of the Departments and other interested personnel within the City. Even when they are familiar with the operations of the departments, they have never worked with this data before nor have they ever seen any data visualization or summary metric.

Design a ~15-20 minutes presentation to effectively communicate the insights, process, and next steps to this audience. Prepare a slide deck, then send it along with your answers to the previous exercises. Many attendees consider themselves non-technical people, but be prepared to answer technical questions too.

Additionally, please be prepared to present this slide deck during your next interview.