The Data Analytics INSTITUTE

Project 1:

AirBnB Investment

PROJECT OVERVIEW

You're working for a client who wishes to invest in an Airbnb property in Washington, D.C. Before your client decides to invest, they'd like clean data about Airbnb performance in D.C.'s neighbourhoods that supports a clear recommendation for an investment in a specific market.

DELIVERABLES AND TIMELINE

1. **Presentation**
2. **Workbook**

1. 10-Minutes Presentation

- **Format**: Google Slides or PDF (export any Keynote or PPT files).
- **Description**:
    - Include all relevant prompts & sub-prompts you chose.
    - Reference any data selected from the original file.
    - Describe any cleaning methods used to remove erroneous data.
    - List recommendations based on your sub-prompt.
    - Present your findings to the class.

2. Excel Workbook

- **Format**: A spreadsheet worksheet containing your calculations.
- **Description**:
    - Include clean listing data with the requested data points.
    - Results of analysis are presented in separate worksheets, formatted, and (if applicable) visualized.
    - Include exploratory efforts using PivotTables, visualizations, and statistical review (if any).
    - Format: Excel file with multiple worksheets:
        - Sheet 1: Clean listing data with the requested data points.
        - Sheet 2: Summary of data cleansing.
        - Other sheets (as utilized): Results of analysis and exploratory efforts.

DATA SET

You have been provided with scraped data captured by a web program with listing information from the Airbnb website. This data may contain unformatted data points with duplicate entries. You will want to clean and format the data prior to performing exploratory analysis — this will help you better understand the available data and build some business context.

PROMPTS

Main prompt: **Should our investor invest in an Airbnb hotel in Washington, D.C.? If so, in which neighbourhood should they invest?**

Choose one of the following sub-prompts to help guide your analysis:

- Prompt 1: Host revenue — How much revenue do successful hosts generate?
- Prompt 2: Property reviews — Which property types receive the most positive reviews?
- Prompt 3: Neighbourhood popularity — Which neighbourhoods host the most listings?
- Prompt 4: Neighbourhood sentiment — Which neighbourhoods receive the most positive reviews?

GETTING STARTED: PROJECT STEPS

1. Project Step 1: Preparing our data set.
    a. Regardless of the sub-prompt you chose, you must have clean data before beginning your analysis. Follow these steps to prepare the Airbnb data set for analysis.
2. Project Step 2: Data exploration and analysis.
    a. Leveraging the clean Airbnb data set, use this step to answer your assigned prompt.
3. Project Step 3: Visualize, summarize, and present.
    a. Polish your work and findings from your data exploration and analysis by distilling your insights with the steps provided.

PROJECT STEP 1: PREPARING OUR DATA SET (ALL PROMPTS)

- Data Cleaning
    - Remove duplicate and erroneous data.
        - Remove listings without any reviews and duplicate rows where the bot may have re-recorded listing data.
        - The “id” (col A) is the ID you should use to find duplicates. The “host_id” (col Q) is the ID of the host, but hosts can have more than one listing.
        - Are descriptions of the property and summaries of the neighbourhood going to help your analysis?
    - Standardize the entry of “State,” “City,” and “Neighbourhood.”
        - “Find and Replace” can be used to find all the values of a specified text value (e.g., “Rd.”) and replace them with a different text value (e.g., “Road”). This is the same as in Microsoft Word or Google Docs, in which you can find and replace words. For example, if you wanted to replace all of the occurrences of the name “John Smith” with the name “Jane Smith,” you could use the “Find and Replace” functionality.
        - Alternatively, you could identify the values in the data and translate them into a standard format of your choice. Then, by adding a new temporary column, you could use VLOOKUP to translate entries to the spelling of your choice.

PROJECT STEP 2: DATA EXPLORATION AND ANALYSIS

(BY PROMPT) PROMPT 1: How much revenue do successful hosts make?

- Estimate revenue per listing (each row is considered a listing).
    - Use the following assumptions:
        - Each booking always has two guests, unless the listing only accommodates one.
        - The booking is always for the minimum number of days allowed.
        - Only half of the bookings generate a review.
        - Column to be used: Column BF, “price:” The price for a one-night stay.

- Step 1: Calculate a proxy “estimated number of stays" for each listing by assuming that 50 percent of customers who stayed left a review (use the data in the “number_of_reviews” column). If 10 customers leave a review, you can assume the listing had 20 stays in total (create a new column with a number that is an estimate for how many stays each listing has received).

- Step 2: Multiply the “Price” by the minimum number of nights to get an “estimated revenue per booking”.

- Step 3: Calculate an “estimated total revenue” for each listing by multiplying the estimated revenue per booking by the estimated number of stays.

- **Format**: Have 3 different columns for these three different steps.

- Build several PivotTables in order to quickly explore the data at a high-level:
  - PivotTables should contain the host name, total revenue, and number of listings (make sure to exclude listings with no bookings).
  - Additional PivotTables are welcome and useful for gaining a better understanding of the data.
  - Format: Currency/ thousands separators should be used where appropriate.

PROMPT 2: Which property types receive the most positive reviews?

- Build several PivotTables in order to quickly explore the data at a high level:
  - PivotTables should contain the property type, number of listings (make sure to exclude listings with no bookings), and average rating.
  - Additional PivotTables are welcome and useful for gaining a better understanding of the data.
  - Format: Currency/ thousands separators should be used where appropriate.

PROMPT 3: Which neighbourhoods host the most listings?

- Build several PivotTables in order to quickly explore the data at a high level:
  - PivotTables should contain the host name and number of listings (make sure to exclude listings with no bookings).
  - Additional PivotTables are welcome and useful for gaining a better understanding of the data.
  - Format: Currency/ thousands separators should be used where appropriate.

PROMPT 4: Which neighbourhoods receive the most positive reviews?

- Build several PivotTables in order to quickly explore the data at a high level:
  - PivotTables should contain the neighbourhood name, number of listings (make sure to exclude listings with no bookings), and average rating.
  - Additional PivotTables are welcome and useful for gaining a better understanding of the data.
  - Format: Currency/ thousands separators should be used where appropriate.

PROJECT STEP 3: VISUALIZE, SUMMARIZE, AND PRESENT (ALL PROMPTS)

- Visualize insightful data for easier comprehension in support of your argument.
  - Format: Copy and paste the results of your PivotTable(s) and adjust them to only show the relevant findings. Create a chart using the format that best fits your findings' needs.
  - Your client is unwilling to go into untested markets — consider limiting the data you visualize to the top 10 best performers.

- Summarize relevant statistics for various data points (max, median, mode).
  - Identify the listing/host that generates the most revenue.
    - xy percent rating with an entire two-bedroom apartment under $200/night.
  - Provide a profile of Airbnb activity in Washington, D.C.
    - xy percent are two-bedroom apartments, cd percent use airbeds, st percent of all listings are in X neighbourhood, etc.
  - Identify the top-performing hosts and listings.
  - Generate list of prospect hosts to target.

- Organize your workbook appropriately.
- Present of your findings to the class:
  - Each student or team will present.
  - Each presentation should last five minutes.
  - In addition to your conclusions, discuss the cleaning techniques you used
  - While others are presenting, take notes on the important points made by each team.

PROJECT STEP 4: RECOMMENDATION (ALL PROMPTS)

- Finally, discuss your overall recommendation. Review the original prompt: "Should our investor invest in an Airbnb hotel in Washington, D.C.? If so, in which neighbourhood should they invest?"
  - **Create a final recommendation** to the investor based on pertinent data from your analysis.
  - Keep in mind that the client wishes to understand the local market and its potential to host a profitable Airbnb property — demand is key, as is the price point.