Final project | Airbnb Listings & Reviews

Total score: 120 points

Introduction

Airbnb data for 1000 listings in 10 major cities, including information about hosts, pricing, location, and room type, along with over 82 thousand historical reviews.

About the dataset

File: Listings_small.csv/Reviews_small.csv

Table	Field	Description
Listings	listing_id	Listing ID
	Name	Listing Name
	host_id	Host ID
	host_since	Date the Host joined Airbnb
	host_location	Location where the Host is based
	host_response_time	Estimate of how long the Host takes to respond
	host_response_rate	Percentage of times the Host responds
	host_acceptance_rate	Percentage of times the Host accepts a booking request
	host_is_superhost	Binary field to determine if the Host is a Superhost
	host_total_listings_count	Total listings the Host has in Airbnb
	host_has_profile_pic	Binary field to determine if the Host has a profile picture
	host_identity_verified	Binary field to determine if the Host has a verified identity
	neighbourhood	Neighborhood the Listing is in
	District	District the Listing is in
	City	City the Listing is in
	Latitude	Listing's latitude
	Longitude	Listing's longitude
	property_type	Type of property for the Listing
	room_type	Type of room type in Airbnb for the Listing
	accommodates	Guests the Listing accomodates
	Bedrooms	Bedrooms in the Listing
	Amenities	Amenities the Listing includes
	Price	Listing price (in each country's currency)
	minimum_nights	Minimum nights per booking

	maximum_nights	Maximum nights per booking
	review_scores_rating	Listing's overall rating (out of 100)
	review_scores_accuracy	Listing's accuracy score based on what's promoted in Airbnb (out of 10)
	review_scores_cleanliness	Listing's cleanliness score (out of 10)
	review_scores_checkin	Listing's check-in experience score (out of 10)
	review_scores_communication	Listing's communication with the Host score (out of 10)
	review_scores_location	Listing's location score within the city (out of 10)
	review_scores_value	Listing's value score relative to its price (out of 10)
	instant_bookable	Binary field to determine if the Listing can be booked instantly
Reviews	listing_id	Listing ID
	review_id	Review ID
	date	Review Date
	reviewer_id	Reviewer ID

Part 1 | Statistics & Excel (20 points)

File: Listing_small.csv & Reviews_small.csv

In this part, you are going to find some statistics about the data with the help of Excel. You are going to need Pivot Tables (2 separate for each file) to make your analysis easier. Some questions need to be calculated using Functions.

- How many reviews per listing name
- What is the average number of minimum nights
- What is the minimum review scores communication
- What is the average price
- What is the standard deviation of price
- Does the price follows Normal Distribution
- Find the correlation between price and review scores rating. Explain the result

Part 2 | SQL (20 points)

In this part, you are going to have access to the Database **Airbnb Listing** in SQL Server Management Studio of CuriousIQ and respond to the following questions:

- You have been requested to extract some information about the Host. Write a query that displays the following columns (ensure that the Name of each column is as follows):
    - Listing Name, Host location, Host Response Time, Host Response Rate, Host is superhost, Host Total listings count, Host Identity verified
- Find the listings with **price** less than 100/night
- What is the minimum **review scores rating** per property type

- How many reviews per **neighbourhood**
- How many reviews per **listing name** and what is the listing name with the **most** reviews?

Part 3 | Python (20 points)

File: listings_smal.csv

To be completed: Final Project, Notebook – To be completed.ipynb (Open it in Google Colab)

In this part you are going to understand your data a little bit more using Pandas library and creating some visualizations.

- Find the data types for each column for listings dataset
- Find the descriptive statistics for listings dataset
- Find how many null values per column
- Find how many listings per property type
- Replace f/t values of instant bookable column to No/Yes respectively **(hint: inplace = True)**
- Sort dataset by price in descending order **(hint: inplace = True)**
- Fill empty (nan) values of host_response_time to N/A
- Find how many listings per host_response_time in a pie chart **(hint: unique() and value_counts())**
- Find the average price per city and visualize it in a sorted, bar chart

Part 4 | Data Visualization (40 points)

File: PowerBI - Final Project (draft).pbix (the data are already imported)

In this part, you are going to create a dashboard in Power BI. After answering all these questions, it's worth to share some insights about what you've find concerning Listings, Hosts and Reviews.

You will be given the questions that you have to answer but the most critical part here is to choose the correct Visualization/Chart and ensure that once you look at the chart you are directly understanding the answer.

In the parentheses, you are provided with the name of the page. Creating a dashboard doesn't mean that it is only one page. Each page has to be categorized in order once you are communicating your insights to have a flow in your speech and do not go to one topic and another.

- (General) Find how many listing names exist
- (General) Find how many listings are instantly bookable
- (General) Display in the best visual way the number of listings per Longitude and Latitude

- (General & Listing) Find how many reviews collected (For all the listings and for one listing)
- (Listing) Display the minimum and maximum nights in a single visual
- (Listing) Add filter for name and property type
- (Listing) Display per Listing Name the following information: Name, Neighbourhood, District, City, Property Type, Room Type, Accommodates, bedrooms, price and review scores rating
- (Listing) Display in the best visual way the location of the Listing (hint: Longitude and Latitude)
- (Property Type) Find how many listings per property type
- (Property Type) Find how many listings per room type
- (Property Type) Find how many listings per bedroom (bedrooms available)
- (Property Type) Find how many accommodates per property type
- (Reviews) Find how many reviews per Year/Quarter
- (Reviews) Find how many reviews per Listing Name (We want to see also what the Listing with the highest number of reviews is)
- (Reviews) Find how many reviews per Reviewer (We want to see also what the Reviewer with the highest number of reviews is)
- (Prices) Find the average listing price
- (Prices) Find average price per City (and find the top valued city – with least average price)
- (Prices) List the price of each Listing Name
- (Prices) Find the average price per Year/Month (host_since)
- (Hosts) Find how many superhosts exist
- (Host) Display in the best visual way the location of the Host with the number of Listings
- (Host) List per host the host response rate

Part 5 | Create a presentation (10 points)

Congratulations for making it! In this part you have to communicate your insights because you really answer a lot of questions about the data.

The presentation must be the simplest one. It must be exactly 5+1 pages (without the cover)! Each page will be based on the Power BI's dashboard pages. E.g. 1st page General, 2nd page Listings, etc. You will write in simple words what is the insight that you made. It can be whatever you feel that is worth saying to someone and is important to take a decision or to understand the market. The last page (6th) will be whatever you would like, maybe something extra.

Part 6 | Upload you project to Github (10 points)

In order for you to showcase your skills as a Data Analyst, you have to share your work and create a portfolio. How should people know that you are good at what you are doing? How can you prove that you learn after all?

This is the most critical step here and you are one step closer to land your first job as a Data Analyst!

- Create an account on Github (Sign Up)
- Log In to your account (Sign In)
- On the upper right corner, next to your profile picture, select **New Repository**
- As a repository name write **data-analysis-curiousiq**
- In **Description**, write the Part 1, 2, 3 and 4 explanations found in this document
- Make it **Public**
- Check **Add a README file**
- **Create Repository**
- Create for each part a folder (1st one: Statistics and Excel, 2nd one: SQL, 3rd one: Python, 4th one: Power BI)
- Drag and drop each file created to the respective folder
- Create a folder named **Dataset**
- Drag and drop the data files used to the **Dataset** folder
- Write on README file the following:

Curious IQ – Data Analytics | Cohort <your cohort number>
Name: <Your name>
Academic Year: <Your academic year e.g. September 2022-March 2023>

Final Project covering topics:
- Descriptive Statistics
- Excel
- SQL
- Python
- Power BI

In the description you are going to find the questions answer per topic.

Good Luck!