Initial Exploratory Data Analysis in Power BI Lab

PART 1: Identify missing data

You can download the course materials from here:

https://github.com/fenago/cts245X/tree/main/EDA/Exercises

Download Power BI Desktop from here:

https://www.microsoft.com/en-us/download/details.aspx?id=58494

Open the Power BI report named `1_1_identify.pbix` from the Exercises folder. If a pop-up window opens asking you to login, you can click *Cancel*.

Load the CSV file `airbnb.csv` from the Datasources folder.

Create a new *Card* visualization showing the **distinct** count of listings and filter for listings without a price.

Make a copy of this card and modify it to be a table named "Number of Listings with a Blank Price".

Explore the counts of missing data in relation to `review_scores_rating` and `city`.

It appears all listings without a price come from one city, what's the name of that city?

Part 2: Descriptive statistics for a variable

Create a new page named "describe"; rename the first page to "missing".

Create a new *Table* visualization with the count of distinct listings.

Add a new column to the table showing the average `price` .

Add three more columns showing the minimum, median, and maximum for `price` .

Repeat these steps to create a separate *Table* for `host_acceptance_rate` .

Based on the values of the average and median, would you suspect the distribution for **price to be...**

- ... right-skewed
- ... left-skewed
- ... normally distributed

Part 3: Imputation for missing data

- Create a new page called "distribution".
- Change the data type of `price` to a decimal number.

Create a new *Table* visualization. It should contain:

- `listing_id` as a distinct count,
- `price` as an average.
- `price` as a median.

Filter the page to show only listings from Sydney.

Create a new *calculated column* called `updated_price` which, if `price` is blank, sets the value to `110` (the median price for Sydney listings), otherwise uses the value of `price` .

You can create a logical statement if there are blank values in a column with the function ISBLANK().

https://learn.microsoft.com/en-us/dax/isblank-function-dax

Add 2 more variables to the table:

- `updated_price` as a median.
- `updated_price` as an average.

Change all values to currency data types with 2 decimals.

Note how this is a way to add 2 different metrics to a table: median and average.

How did imputing the blank with the median effect the average listing price in Sydney?

- ○ Increased by $0.51
- ○ Decreased by $0.51
- ○ Stayed the same.