Business Data Analysis
Integrative Project: Comparative Real Estate Market Analysis

1. Project Objective

The goal of this project is to apply the statistical concepts and methods learned in class to conduct a real-world data analysis. You will collect, organize, analyze, and present data on the residential real estate markets in two different Canadian municipalities.

Your final analysis will compare these two markets, drawing statistically-supported conclusions about their differences and characteristics. This project will test your mastery of data collection, descriptive statistics, inferential statistics, and regression analysis using Microsoft Excel.

2. Project Deliverables

You will submit two files:

1. **A Professional PowerPoint Presentation:** A 10-15 slide presentation that clearly communicates your methodology, key findings, and conclusions. This should be aimed at a business audience.
2. **An Excel Workbook:** A single, well-organized Excel file containing all your raw data, calculations, tables, and graphs. This file must show your work clearly.

3. Part 1: Data Collection

Your first task is to gather your own data set.

1. **Select Your Municipalities:** Choose any two (2) Canadian municipalities that you wish to compare. (e.g., Beaconsfield, QC vs. Kirkland, QC; or a broader comparison like Vaughan, ON vs. Surrey, BC).
2. **Data Source:** Use the website https://www.realtor.ca/.
3. **Sampling:** For *each* of your two chosen municipalities, you must collect a random sample of **50 residential listings** (for a total of n=100 listings).
    - *Tip for random sampling:* Do not just pick the first 50 listings you see. A good method is to go page by page and use a random number generator (=RANDBETWEEN(1, 200) in Excel) to decide which listing to select from that page.
4. **Variables to Collect:** For each of the 100 listings, you must record the following variables in an Excel spreadsheet:
    - **Municipality:** (e.g., "Vaughan")
    - **Listing Price:** (Quantitative)
    - **Property Type:** (Categorical, e.g., "Detached," "Condo/Apartment," "Townhouse")
    - **Number of Bedrooms:** (Quantitative, Discrete)
    - **Number of Bathrooms:** (Quantitative, Discrete)

4. Part 2: Data Analysis & Computations (in Excel)

This is the core of your project. You must perform the following analyses for *each* municipality and for *comparison*.

A. Descriptive Statistics

- **Measures of Location:** Calculate the Mean for Listing Price.
- **Measures of Variation:** Calculate the Variance, and Standard Deviation for Listing Price.
- **Frequency Distribution:** Create a frequency distribution and a **Histogram** for the Listing Price of each municipality.
- **Charts:** Create a **Bar Chart** or **Pie Chart** that shows the breakdown of Property Type (e.g., % Detached, % Condo) for each municipality. A stacked or side-by-side bar chart is excellent for comparison.
- **Box Plots:** Compute the values necessary for creating box-and-whisker plots for both municipalities. Draw the box-and-whisker plots side-by-side on the same axis.

B. Probability

- Based on your sample data, answer the following (show your calculations):
  1. What is the probability that a randomly selected home in Municipality A costs over $800,000? (Repeat for Municipality B).
  2. What is the probability that a randomly selected home in Municipality A is "Detached" *OR* has "3 Bedrooms"?

C. Inferential Statistics: Estimation

- For *each* municipality, calculate the **95% confidence interval** for the true mean Listing Price.
- Clearly interpret what this confidence interval means in the context of your data.

D. Inferential Statistics: Hypothesis Testing

- This is your main comparison. You will perform a hypothesis test to determine if there is a statistically significant difference in the mean listing prices between the two municipalities.
  1. **State Hypotheses:** Clearly state your Null Hypothesis (H_0) and Alternative Hypothesis (H_a) for a two-tailed test.
  2. **Set Significance Level:** Use alpha = 0.05.
  3. **Perform Test:** Conduct the appropriate t-test for two independent samples. (You should first test if the variances are equal or unequal and use the correct test).
  4. **Find p-value:** Report the p-value from your test.
  5. **Conclusion:** State your decision (Reject or Fail to Reject H_0) and write a clear conclusion in plain English: "Based on my sample, there (is / is not) a significant difference in the mean home prices between..."

E. Correlation & Regression

- For this section, you can analyze each municipality separately or combine them into one

large dataset (specify which you are doing).

1. **Correlation:** Calculate the correlation coefficient (r) between:
   - Listing Price and Number of Bedrooms
   - Listing Price and Number of Bathrooms
   - Interpret the strength and direction of these relationships.
2. **Simple Linear Regression:** Develop a simple linear regression model where Listing Price is the dependent variable (Y) and Number of Bedrooms is the independent variable (X).
3. **State the Equation:** Provide the final regression equation (e.g., Price = b + m . # Bedrooms).
4. **Interpret Coefficients:** Clearly explain what the b coefficient (slope) means in this context.
5. **R-Squared:** Report and interpret the R^2 value.

5. Part 3: Professional Presentation (Deliverable 1)

Your presentation should "tell the story" of your data. Do not just paste screenshots of Excel. Re-create your key charts and tables in a clean, professional format.

Suggested Structure (10-15 slides):

1. **Title Slide:** Project Title, Your Name, Course.
2. **Introduction:** What two municipalities did you choose? What was your objective?
3. **Methodology:** Briefly explain *how* you collected your data (source, sample size).
4. **Descriptive Statistics:** Show your key descriptive stats (mean, median, std. dev.). Include your histograms and property type charts. *Interpret* what they show.
5. **Inferential Statistics:**
   - Present your 95% confidence intervals and explain what they mean.
   - Walk through your hypothesis test (hypotheses, p-value, and final conclusion).
6. **Correlation & Regression:**
   - Present your correlation findings.
   - Show your regression equation and its R^2 value. Explain what it tells you about the market.
7. **Conclusion & Limitations:**
   - Summarize your single most important finding.
   - Briefly mention any limitations of your study (e.g., small sample size, other variables not included).

6. Part 4: Excel Workbook (Deliverable 2)

Your Excel file must be organized and easy for your professor to audit. Use separate tabs for each part of the analysis.

- **Tab 1: Raw Data:** Your full dataset of 100 listings.
- **Tab 2: Descriptive Stats:** All calculations for Part 2A (means, medians, std. devs) and the tables/charts.

- **Tab 3: Probability:** Calculations for Part 2B.
- **Tab 4: Inferential Stats:** All calculations for your confidence intervals and t-test (Part 2C & 2D).
- **Tab 5: Regression:** Your correlation matrix and regression analysis output (Part 2E).

All calculations must be done using Excel formulas. Do not just type in numbers.

Due Date: [December 17th, 2025]