AN ANALYSIS OF HOUSING PRICE & REAL ESTATE IN 2023

Goal:

The goal of the project is to study the Canadian housing market in 2023 to offer information on potential investment opportunities in different regions and types of properties. Using data analysis methods and visual representations, the project aims to grasp trends in housing prices, the impact of demographics.

Datasets Used & Licensing References:

1. Housing Price & Real Estate - 2023:
   - Data Source: Kaggle
   - Licensing: Public Domain
   - URL: https://www.kaggle.com/datasets/reenapinto/housing-price-and-real-estate-2023
   - Description: This dataset provides information about housing prices and real estate tr ends in Canada for the year 2023. It includes data such as property prices, descriptions, locations, and other relevant attributes.

2. Community Points:
   - Data Source: Calgary Open Data
   - Licensing: Open Government License - City of Calgary
   - URL: https://data.calgary.ca/Base-Maps/Community-Points/j9ps-fyst/data
   - Description: This dataset contains community points in Calgary, where each point identifies the centroid of a specific community. It includes geographic coordinates and other community-related attributes.

Citations:

- Kaggle Housing Price & Real Estate - 2023: Reena (Owner), Kaggle
- Calgary Open Data - Community Points: The City of Calgary, Calgary Open Data

By utilizing these datasets and adhering to the respective licensing terms, the project aims to provide comprehensive insights into the Canadian housing market in 2023, facilitating informed decision-making for investors.

Introduction:

- In this study, we analyze the Canadian housing market in 2023 using data from reputable sources like Kaggle and Calgary Open Data.
- Our objective is to provide actionable insights for stakeholders.
- We follow the CRISP-DM methodology, beginning with a thorough understanding of the business context and data.
- We formulate relevant business questions and create visualizations to address them effectively.
- Our findings are presented through a Power BI report, offering a comprehensive view of housing trends and investment opportunities.

1. Business Understanding

Objective:

- The objective is to analyze trends in the housing market, particularly focusing on information on potential investment opportunities in different regions and types of properties.

Background:

- Housing market experts believe there's a shift towards a seller's market due to increased demand and limited supply.
- Factors contributing to increased demand include millennials entering the market as first-time buyers.
- Both residential and commercial sectors are experiencing heightened demand.
- Some experts speculate that this demand surge might be contributing to a housing crisis.
- Canada is implementing measures like capping immigration to address potential housing issues.

Stakeholders:

- Real estate agencies, Homeowners, Prospective homebuyers, Property investors

Success Criteria:

- Ability to provide insights into factors driving the market (e.g., demographics, economic trends, housing supply).

Constraints:

- Availability and quality of data: Ensuring access to reliable data sources for analysis.
- Regulatory factors: Compliance with real estate regulations and laws.

Assumptions:

- The analysis is based on 2023 housing prices listed on various websites, trends and may change depending on future economic and social developments.

Risks and Contingencies:

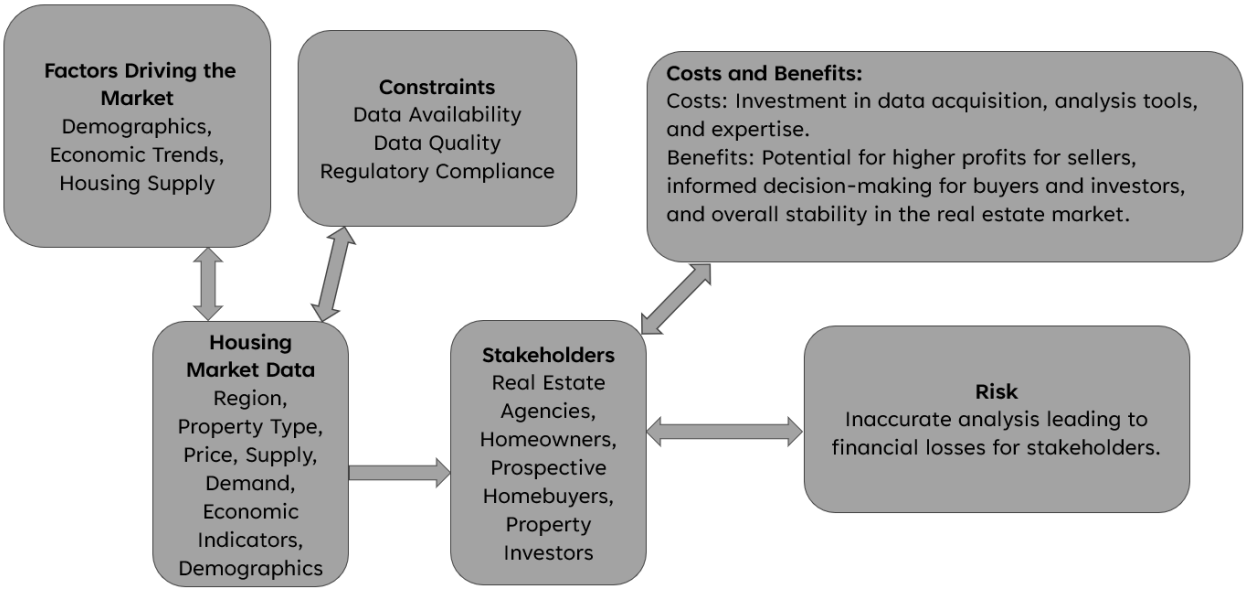
- Risk: Inaccurate analysis leading to financial losses for stakeholders.

- Contingency: Regular monitoring and updating of models based on new data and market changes.

Costs and Benefits:

- Costs: Investment in data acquisition, analysis tools, and expertise.
- Benefits: Potential for higher profits for sellers, informed decision-making for buyers and investors, and overall stability in the real estate market.

Conceptual Model (Data Map):



Data Dictionary:

Topic	Definition
Region	Refers to a geographical area within Canada, such as provinces, cities, or neighborhoods, where housing market data is collected and analyzed.
Property Type	Describes the category or classification of a property, such as single-family homes, condominiums, or commercial properties, which influences its market value and demand.
Demand	Refers to the level of interest or desire for properties in a specific region or property type, influenced by factors such as population growth, economic conditions, and buyer preferences.
Supply	Indicates the availability of properties for sale or rent within a given region or property type, impacting market dynamics, pricing, and competition among buyers and sellers.
Regulatory Compliance	Involves adherence to real estate regulations, laws, and ethical standards governing data collection, analysis, and reporting to ensure legality, transparency, and ethical conduct in the housing market.
Demographics	Encompasses the statistical characteristics of a population within a region, including age, income, household composition, and cultural diversity, which influence housing demand and market trends.

2. **Data Understanding**

- Housing Price & Real Estate 2023: Contains data on house listings, including price, geographic location, property type, and additional features influencing housing value.

```
# Import the necessary library and load the datasets
import pandas as pd

# Load Homes for Sale and Real Estate data set
Housing = pd.read_excel("Homes_for_Sale_and_Real_Estate.xlsx")
Housing.head(5)
```

✓ 0.3s

	Address	Price	Description	Place	Beds	Bath	Sq.Ft	Website
0	3704 42 St SW	979999	CA AB T3E 3N1	Glenbrook	4	3.5	1813	Century 21 Bravo Realty
1	30 Mahogany Mews SE #415	439900	CA AB T3M 3H4	Mahogany	2	2.0	1029	Century 21 Bamber Realty Ltd.
2	273 Auburn Shores Way SE	950000	CA AB T3M 2E9	Auburn Bay	4	2.5	2545	Exp Realty
3	235 15 Ave SW #404	280000	CA AB T2R 0P6	Beltline	2	2.0	898	RE/MAX Realty Professionals
4	24 Hemlock Crescent SW #2308	649000	CA AB T3C 2Z1	Spruce Cliff	2	2.0	1482	Charles

```
Housing.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3360 entries, 0 to 3359
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Address      3360 non-null   object
 1   Price        3360 non-null   int64
 2   Description  3360 non-null   object
 3   Place        3344 non-null   object
 4   Beds         3360 non-null   int64
 5   Bath         3360 non-null   float64
 6   Sq.Ft        3360 non-null   int64
 7   Website      3359 non-null   object
dtypes: float64(1), int64(3), object(4)
memory usage: 210.1+ KB
```

3 & 4. Data Preparation: Cleaning and Modeling:

Clean and preprocess the datasets to handle missing values, outliers, and inconsistencies.

1. Checked blank rows and printed them. There were empty values in the Place and Website columns. As our analysis relied on the Place data, it was necessary to retain these rows.

```
# Check for blank rows
blank_rows_housing = Housing[Housing.isnull().any(axis=1)]

# Display the blank rows, if any
if not blank_rows_housing.empty:
    print("Blank rows found in Housing DataFrame:")
    print(blank_rows_housing)
else:
    print("No blank rows found in Housing DataFrame.")
```

```
Blank rows found in Housing DataFrame:
                    Address    Price     Description   Place  Beds  Bath  \
122   1066 Creekside Blvd SW   580900   CA  AB T2X5K6     NaN     1   1.5
160     341 Walcrest View SE   820000  CA  AB T2X 4V9  Walden     5   3.5
641     62 Royston Terrace NW  849900  CA  AB T3L 0J2     NaN     3   2.5
1154          3250 84 St SE  2400000  CA  AB T2B 3C1     NaN     6   2.0
1174     8535 19 Ave SE #424   455000  CA  AB T2A 7W8     NaN     2   1.5
1245      99 Taralake Way NE   672000   CA  AB T3J0A7     NaN     5   3.5
1324       148 Savanna Dr NE   850000   CA  AB T3J2H5     NaN     4   3.0
1519            9110 34 Ave  2299000  CA  AB T1X 0L5     NaN     6   4.5
1630   71 Lynx Meadows Dr NW  2000000  CA  AB T3L 3L9     NaN     6   4.5
1846     8535 19 Ave SE #421   505000  CA  AB T2A 7W8     NaN     3   2.5
1867          4520 84 Ave NE  1000000  CA  AB T3J 4C4     NaN     5   4.0
2019      99 Royston Rise NW   794900  CA  AB T3L 0J2     NaN     3   2.5
2361  1161 Creekside Blvd SW   667500   CA  AB T2X5K5     NaN     3   2.5
2629     8535 19 Ave SE #417   488900  CA  AB T2A 7W8     NaN     4   3.5
3033       30 Forzani Way NW  2388000  CA  AB T3Z 1L5     NaN     5   3.5
3311     902 Bluerock Way SW   702500  CA  AB T2Y 0S5     NaN     3   2.5
3356         4111 162 Ave SW  8000000  CA  AB T2Y 0N7     NaN     5   4.5

      Sq.Ft                          Website
122     964              Maxwell Canyon Creek
160    2235                               NaN
641    2467                         Cir Realty
1154   2147                       RE/MAX Key
...
2629   1365                         Exp Realty
3033   2554                         Cir Realty
3311   2096                              Bode
3356   9031  Diamond Realty & Associates Ltd.
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

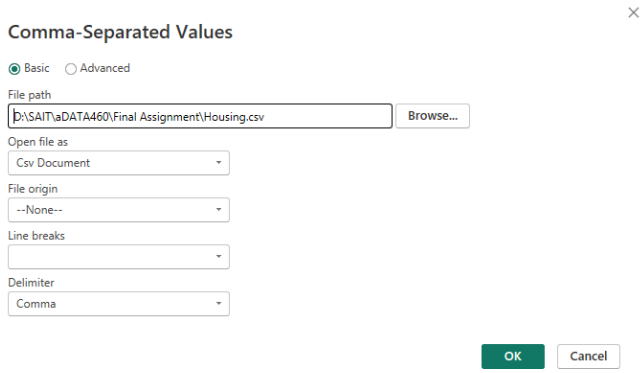
2. To preserve the rows, I decided to replace them with text extracted from the address column.

```python
#%%

import re

# Function to extract text from address
def extract_text(address):
    return re.sub(r'\d+', '', address).strip()

# Apply the function to create a new column with only text from 'Address'
Housing['TextAddress'] = Housing['Address'].apply(extract_text)

# Store rows where 'Place' is blank before replacement
blank_rows_before = Housing[Housing['Place'].isnull()]

# Replace blank rows in 'Place' column with text from 'TextAddress' column
Housing['Place'] = Housing['Place'].fillna(Housing['TextAddress'])

# Store rows where 'Place' was replaced
replaced_rows = Housing.loc[blank_rows_before.index]

# Drop the 'TextAddress' column if you don't need it anymore
Housing.drop(columns=['TextAddress'], inplace=True)

# Print replaced rows
print("Replaced Rows:")
print(replaced_rows)
```

```
Replaced Rows:
                   Address    Price     Description              Place  \
122   1066 Creekside Blvd SW   580900   CA  AB T2X5K6   Creekside Blvd SW
641    62 Royston Terrace NW   849900  CA  AB T3L 0J2  Royston Terrace NW
1154          3250 84 St SE  2400000  CA  AB T2B 3C1              St SE
1174     8535 19 Ave SE #424   455000  CA  AB T2A 7W8           Ave SE #
1245      99 Taralake Way NE   672000   CA  AB T3J0A7    Taralake Way NE
1324       148 Savanna Dr NE   850000   CA  AB T3J2H5      Savanna Dr NE
1519            9110 34 Ave  2299000  CA  AB T1X 0L5               Ave
1630   71 Lynx Meadows Dr NW  2000000  CA  AB T3L 3L9  Lynx Meadows Dr NW
1846     8535 19 Ave SE #421   505000  CA  AB T2A 7W8           Ave SE #
1867          4520 84 Ave NE  1000000  CA  AB T3J 4C4             Ave NE
2019      99 Royston Rise NW   794900  CA  AB T3L 0J2    Royston Rise NW
2361  1161 Creekside Blvd SW   667500   CA  AB T2X5K5   Creekside Blvd SW
2629     8535 19 Ave SE #417   488900  CA  AB T2A 7W8           Ave SE #
3033       30 Forzani Way NW  2388000  CA  AB T3Z 1L5     Forzani Way NW
3311     902 Bluerock Way SW   702500  CA  AB T2Y 0S5     Bluerock Way SW
3356         4111 162 Ave SW  8000000  CA  AB T2Y 0N7             Ave SW
```

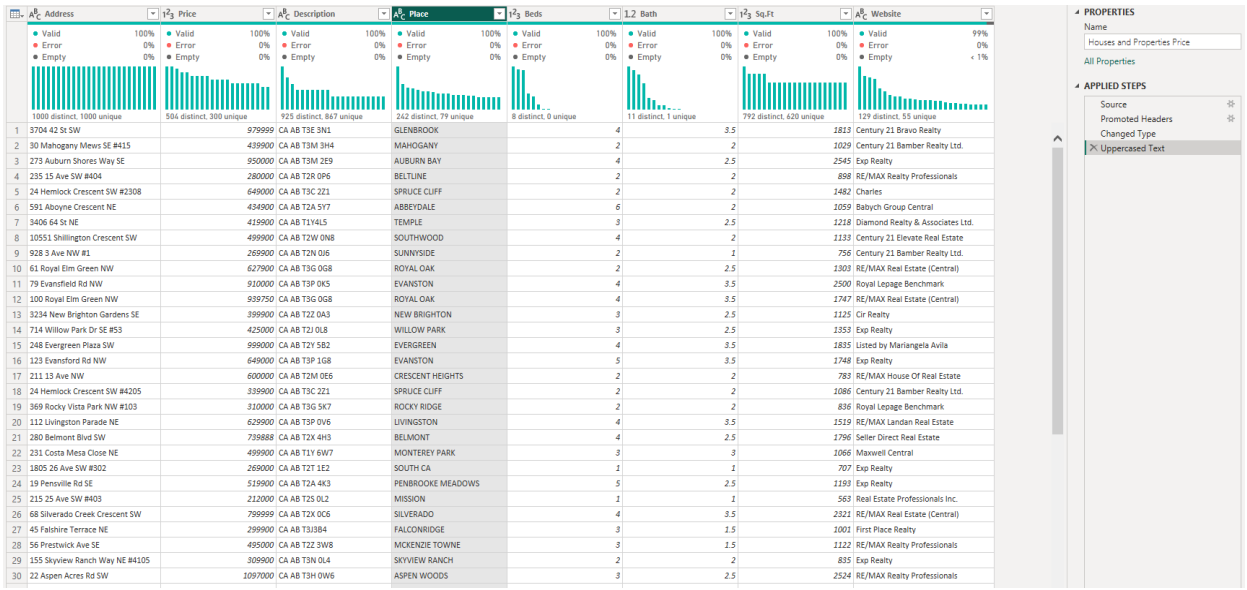
3. Import Data

   Then, we imported data into power bi using import data from csv option.



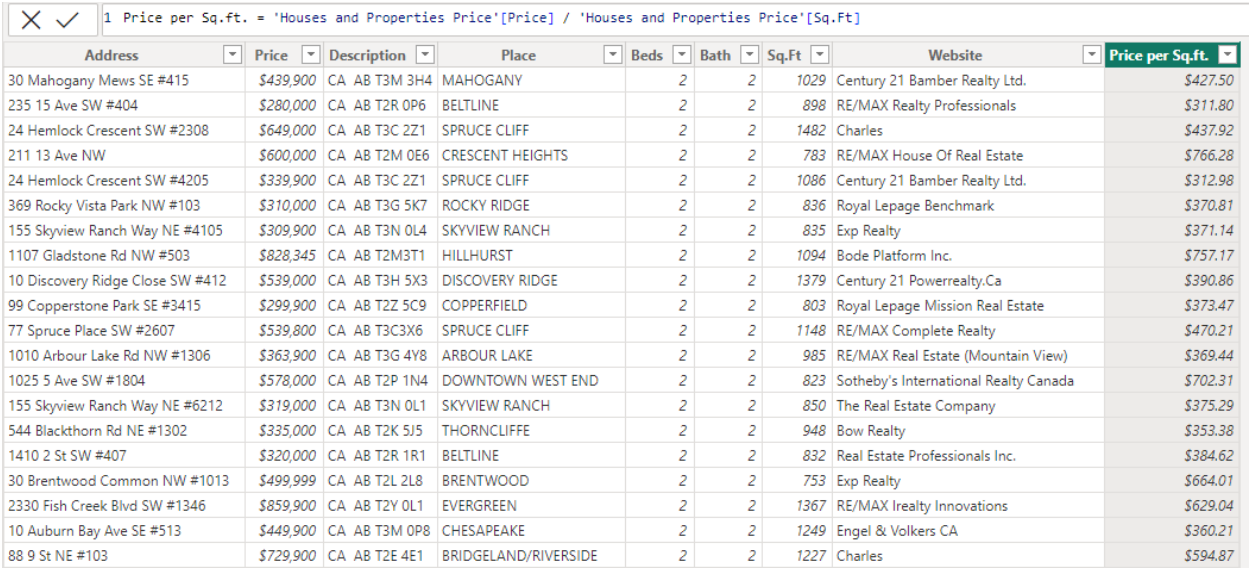
4. Cleaning the data and transforming the data

- Checked to see if we have any unnecessary columns since all columns were relevant in our case, we decided to keep all the columns.
- Correcting Errors: There were no errors to fix when we loaded the data.
- We made sure the data types were appropriate for the data set and columns were named appropriately.



	Address	Price	Description	Place	Beds	Bath	Sq.Ft	Website
1	3704 42 St SW	979999	CA AB T3E 3N1	GLENBROOK	4	3.5	1813	Century 21 Bravo Realty
2	30 Mahogany Mews SE #415	439900	CA AB T3M 3H4	MAHOGANY	2	2	1029	Century 21 Bamber Realty Ltd.
3	273 Auburn Shores Way SE	950000	CA AB T3M 2E9	AUBURN BAY	4	2.5	2545	Exp Realty
4	235 15 Ave SW #404	280000	CA AB T2R 0P6	BELTLINE	2	2	898	RE/MAX Realty Professionals
5	24 Hemlock Crescent SW #2308	649000	CA AB T3C 2Z1	SPRUCE CLIFF	2	2	1482	Charles
6	591 Aboyne Crescent NE	434900	CA AB T2A 5Y7	ABBEYDALE	6	2	1059	Babych Group Central
7	3406 64 St NE	419900	CA AB T1Y4L5	TEMPLE	3	2.5	1218	Diamond Realty & Associates Ltd.
8	10551 Shillington Crescent SW	499900	CA AB T2W 0N8	SOUTHWOOD	4	2	1133	Century 21 Elevate Real Estate
9	928 3 Ave NW #1	269900	CA AB T2N 0J6	SUNNYSIDE	2	1	756	Century 21 Bamber Realty Ltd.
10	61 Royal Elm Green NW	627900	CA AB T3G 0G8	ROYAL OAK	2	2.5	1303	RE/MAX Real Estate (Central)
11	79 Evansfield Rd NW	910000	CA AB T3P 0K5	EVANSTON	4	3.5	2500	Royal Lepage Benchmark
12	100 Royal Elm Green NW	939750	CA AB T3G 0G8	ROYAL OAK	4	3.5	1747	RE/MAX Real Estate (Central)
13	3234 New Brighton Gardens SE	399900	CA AB T2Z 0A3	NEW BRIGHTON	3	2.5	1125	Cir Realty
14	714 Willow Park Dr SE #53	425000	CA AB T2J 0L8	WILLOW PARK	3	2.5	1353	Exp Realty
15	248 Evergreen Plaza SW	999000	CA AB T2Y 5B2	EVERGREEN	4	3.5	1835	Listed by Mariangela Avila
16	123 Evansford Rd NW	649000	CA AB T3P 1G8	EVANSTON	5	3.5	1748	Exp Realty
17	211 13 Ave NW	600000	CA AB T2M 0E6	CRESCENT HEIGHTS	2	2	783	RE/MAX House Of Real Estate
18	24 Hemlock Crescent SW #4205	339900	CA AB T3C 2Z1	SPRUCE CLIFF	2	2	1086	Century 21 Bamber Realty Ltd.
19	369 Rocky Vista Park NW #103	310000	CA AB T3G 5K7	ROCKY RIDGE	2	2	836	Royal Lepage Benchmark
20	112 Livingston Parade NE	629900	CA AB T3P 0V6	LIVINGSTON	4	3.5	1519	RE/MAX Landan Real Estate
21	280 Belmont Blvd SW	739888	CA AB T2X 4H3	BELMONT	4	2.5	1796	Seller Direct Real Estate
22	231 Costa Mesa Close NE	499900	CA AB T1Y 6W7	MONTEREY PARK	3	3	1066	Maxwell Central
23	1805 26 Ave SW #302	269000	CA AB T2T 1E2	SOUTH CA	1	1	707	Exp Realty
24	19 Pensville Rd SE	519900	CA AB T2A 4K3	PENBROOKE MEADOWS	5	2.5	1193	Exp Realty
25	215 25 Ave SW #403	212000	CA AB T2S 0L2	MISSION	1	1	563	Real Estate Professionals Inc.
26	68 Silverado Creek Crescent SW	799999	CA AB T2X 0C6	SILVERADO	4	3.5	2321	RE/MAX Real Estate (Central)
27	45 Falshire Terrace NE	299900	CA AB T3J3B4	FALCONRIDGE	3	1.5	1001	First Place Realty
28	56 Prestwick Ave SE	495000	CA AB T2Z 3W8	MCKENZIE TOWNE	3	1.5	1122	RE/MAX Realty Professionals
29	155 Skyview Ranch Way NE #4105	309900	CA AB T3N 0L4	SKYVIEW RANCH	2	2	835	Exp Realty
30	22 Aspen Acres Rd SW	1097000	CA AB T3H 0W6	ASPEN WOODS	3	2.5	2524	RE/MAX Realty Professionals

5. We created a new column with price per square foot.



```
1 Price per Sq.ft. = 'Houses and Properties Price'[Price] / 'Houses and Properties Price'[Sq.Ft]
```

Address	Price	Description	Place	Beds	Bath	Sq.Ft	Website	Price per Sq.ft.
30 Mahogany Mews SE #415	$439,900	CA AB T3M 3H4	MAHOGANY	2	2	1029	Century 21 Bamber Realty Ltd.	$427.50
235 15 Ave SW #404	$280,000	CA AB T2R 0P6	BELTLINE	2	2	898	RE/MAX Realty Professionals	$311.80
24 Hemlock Crescent SW #2308	$649,000	CA AB T3C 2Z1	SPRUCE CLIFF	2	2	1482	Charles	$437.92
211 13 Ave NW	$600,000	CA AB T2M 0E6	CRESCENT HEIGHTS	2	2	783	RE/MAX House Of Real Estate	$766.28
24 Hemlock Crescent SW #4205	$339,900	CA AB T3C 2Z1	SPRUCE CLIFF	2	2	1086	Century 21 Bamber Realty Ltd.	$312.98
369 Rocky Vista Park NW #103	$310,000	CA AB T3G 5K7	ROCKY RIDGE	2	2	836	Royal Lepage Benchmark	$370.81
155 Skyview Ranch Way NE #4105	$309,900	CA AB T3N 0L4	SKYVIEW RANCH	2	2	835	Exp Realty	$371.14
1107 Gladstone Rd NW #503	$828,345	CA AB T2M3T1	HILLHURST	2	2	1094	Bode Platform Inc.	$757.17
10 Discovery Ridge Close SW #412	$539,000	CA AB T3H 5X3	DISCOVERY RIDGE	2	2	1379	Century 21 Powerrealty.Ca	$390.86
99 Copperstone Park SE #3415	$299,900	CA AB T2Z 5C9	COPPERFIELD	2	2	803	Royal Lepage Mission Real Estate	$373.47
77 Spruce Place SW #2607	$539,800	CA AB T3C3X6	SPRUCE CLIFF	2	2	1148	RE/MAX Complete Realty	$470.21
1010 Arbour Lake Rd NW #1306	$363,900	CA AB T3G 4Y8	ARBOUR LAKE	2	2	985	RE/MAX Real Estate (Mountain View)	$369.44
1025 5 Ave SW #1804	$578,000	CA AB T2P 1N4	DOWNTOWN WEST END	2	2	823	Sotheby's International Realty Canada	$702.31
155 Skyview Ranch Way NE #6212	$319,000	CA AB T3N 0L1	SKYVIEW RANCH	2	2	850	The Real Estate Company	$375.29
544 Blackthorn Rd NE #1302	$335,000	CA AB T2K 5J5	THORNCLIFFE	2	2	948	Bow Realty	$353.38
1410 2 St SW #407	$320,000	CA AB T2R 1R1	BELTLINE	2	2	832	Real Estate Professionals Inc.	$384.62
30 Brentwood Common NW #1013	$499,999	CA AB T2L 2L8	BRENTWOOD	2	2	753	Exp Realty	$664.01
2330 Fish Creek Blvd SW #1346	$859,900	CA AB T2Y 0L1	EVERGREEN	2	2	1367	RE/MAX Irealty Innovations	$629.04
10 Auburn Bay Ave SE #513	$449,900	CA AB T3M 0P8	CHESAPEAKE	2	2	1249	Engel & Volkers CA	$360.21
88 9 St NE #103	$729,900	CA AB T2E 4E1	BRIDGELAND/RIVERSIDE	2	2	1227	Charles	$594.87

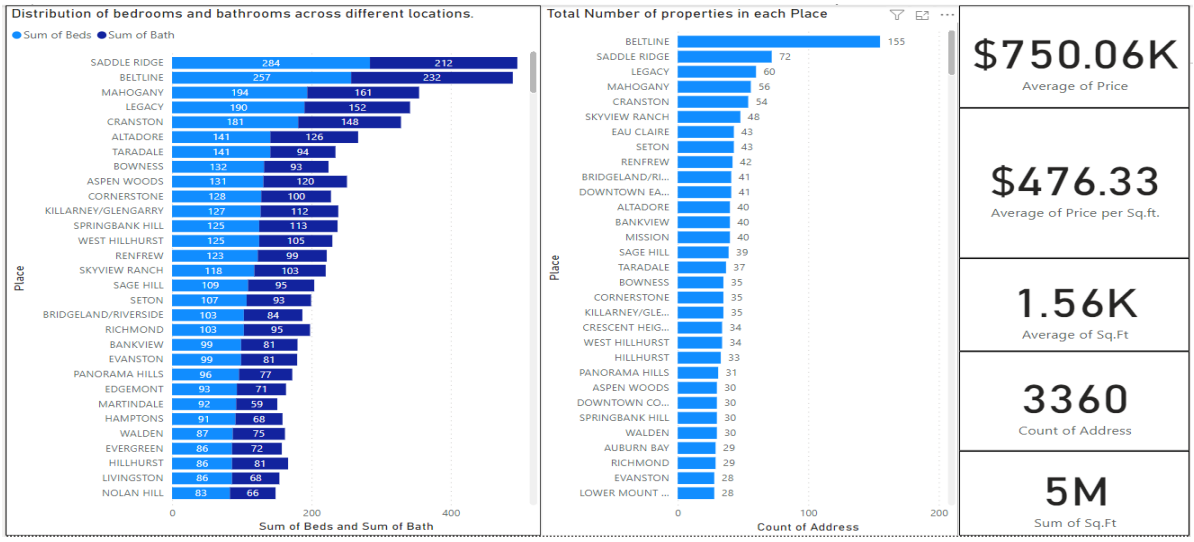
We then changed the formatting of Price and Price per Sq. ft. to currency with 2 decimal places.

5. Evaluation:

Before proceeding to deployment, we conducted a thorough review of the entire process. We did initial analysis on the data set which can be seen below,

Initial Analysis:

From the below Dashboard we can see the Distribution of bedrooms and bathrooms across different locations, Total Number of properties in each place. Average of Price, Average of Price per Sq. ft. Average of Sq. Ft and total number of properties listed and total number of Sq. Ft which is 5M if u convert that in square kilometer it's just approx. 0.5 km2 and total land area of Canada is 825.29 km2 which is not even 1 percent of total land area.



Now let's outline the next steps, since our current dataset lacks geographical data for map visualization, as well as information on sectors and community types, we need to revisit the Business Understanding phase. We will utilize the second dataset and reiterate the CRISP-DM process to address these gaps.

6. Business Understanding

Objective: The goal of the project is to study the Canadian housing market in 2023 to offer information on potential investment opportunities in different regions and types of properties. Using data analysis methods and visual representations, the project aims to grasp trends in housing prices, the impact of demographics.

7. Data Understanding

- Community Points: Provides geographic and demographic data for communities, including centroids for precise location mapping, community structure, and developmental stage.

```
# Load Community_Points data set
Community = pd.read_csv("Community_Points.csv")
Community.head(5)
```

✓ 0.0s

	CLASS	CLASS_CODE	COMM_CODE	NAME	SECTOR	SRG	COMM_STRUCTURE	longitude	latitude	POINT
0	Residential	1	BED	BEDDINGTON HEIGHTS	NORTH	ESTABLISHED	1960s/1970s	-114.085021	51.131633	POINT (-114.08502139544244 51.13163280873361)
1	Residential	1	EVN	EVANSTON	NORTH	COMPLETE	2010s	-114.112453	51.171095	POINT (-114.11245260749 49 51.17109493109596)
2	Residential	1	KIL	KILLARNEY/GLENGARRY	CENTRE	ESTABLISHED	1950s	-114.131727	51.031548	POINT (-114.13172726984385 51.031548429038665)
3	Residential	1	BRA	BRAESIDE	SOUTH	ESTABLISHED	1960s/1970s	-114.106366	50.955993	POINT (-114.10636591786145 50.955992888964275)
4	Residential	1	BLM	BELMONT	SOUTH	DEVELOPING	BUILDING OUT	-114.055252	50.868684	POINT (-114.05525174825 2 50.86868365691495)

```
✓ Community.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 312 entries, 0 to 311
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   CLASS           312 non-null    object
 1   CLASS_CODE      312 non-null    int64
 2   COMM_CODE       312 non-null    object
 3   NAME            312 non-null    object
 4   SECTOR          312 non-null    object
 5   SRG             251 non-null    object
 6   COMM_STRUCTURE  310 non-null    object
 7   longitude       312 non-null    float64
 8   latitude        312 non-null    float64
 9   POINT           312 non-null    object
dtypes: float64(2), int64(1), object(7)
memory usage: 24.5+ KB
```

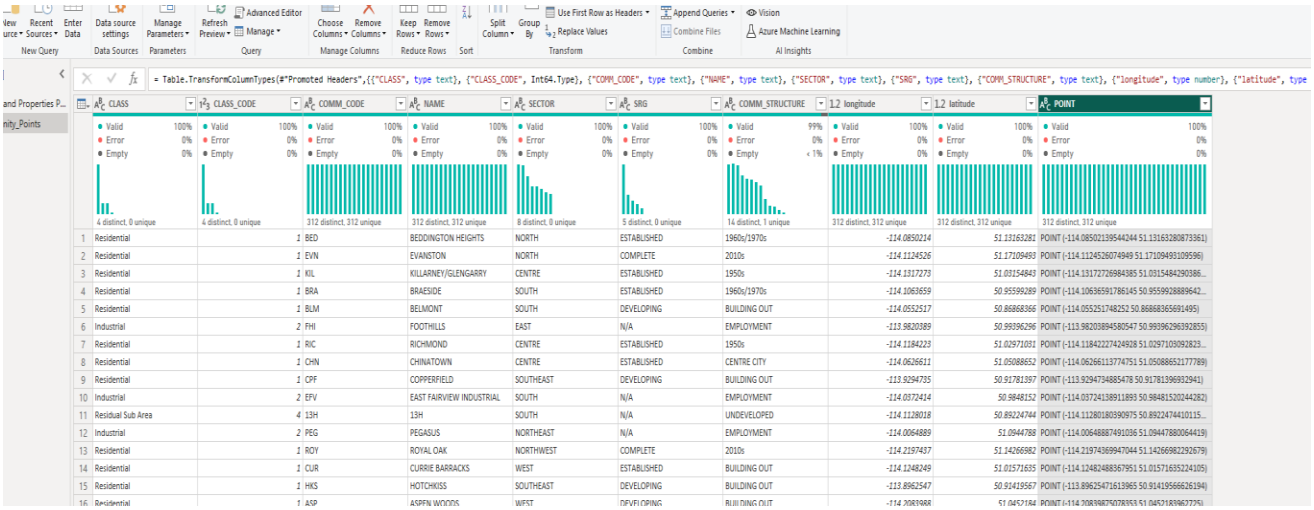
8 &9. Data Preparation: Cleaning and Modeling:

1. We needed latitude and longitude to work with maps, so we downloaded the Community points dataset from source 2 and checked for blank rows.

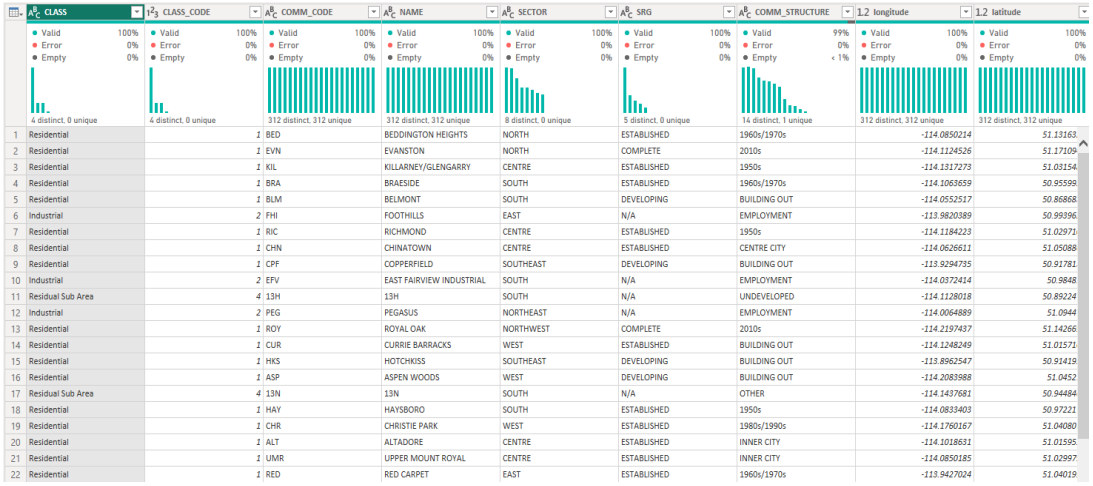
```
# Check for blank rows
blank_rows_community = Community[Community.isnull().any(axis=1)]

# Display the blank rows, if any
if not blank_rows_community.empty:
    print("Blank rows found in Housing DataFrame:")
    print(blank_rows_community)
else:
    print("No blank rows found in Housing DataFrame.")
```

```
Blank rows found in Housing DataFrame:
                CLASS  CLASS_CODE COMM_CODE                      NAME  \
5          Industrial           2       FHI                 FOOTHILLS
9          Industrial           2       EFV  EAST FAIRVIEW INDUSTRIAL
10   Residual Sub Area           4       13H                       13H
11         Industrial           2       PEG                   PEGASUS
16   Residual Sub Area           4       13N                       13N
..                 ...         ...       ...                       ...
290        Industrial           2       MNI     MANCHESTER INDUSTRIAL
296        Industrial           2       HIF                 HIGHFIELD
299  Residual Sub Area           4       13I                       13I
300  Residual Sub Area           4       11B                       11B
309        Major Park           3       FPK           FISH CREEK PARK

       SECTOR  SRG COMM_STRUCTURE   longitude   latitude  \
5        EAST  NaN      EMPLOYMENT -113.982039  50.993963
9       SOUTH  NaN      EMPLOYMENT -114.037241  50.984815
10      SOUTH  NaN    UNDEVELOPED -114.112802  50.892247
11  NORTHEAST  NaN      EMPLOYMENT -114.006489  51.094479
16      SOUTH  NaN           OTHER -114.143768  50.944846
..        ...  ...             ...         ...        ...
290    CENTRE  NaN      EMPLOYMENT -114.057225  51.013261
296    CENTRE  NaN      EMPLOYMENT -114.037201  51.018532
299     SOUTH  NaN    UNDEVELOPED -114.099428  50.892750
300     SOUTH  NaN    UNDEVELOPED -114.164420  50.986652
...
300  POINT (-114.16441971345019 50.986651529493514)
309  POINT (-114.02652092583433 50.909602044237715)

[62 rows x 10 columns]
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

We were using community data points data set for sectors and community data type, so we needed to keep blank rows as we were getting accurate results without it, we created a csv file and then imported it to power bi keeping the blank rows.

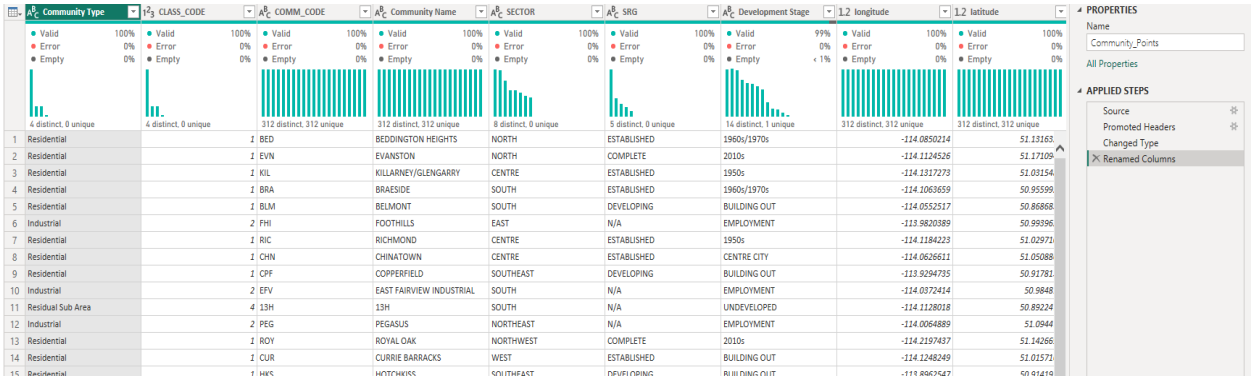


2. We checked Data types and made sure they are appropriate.



	CLASS	CLASS_CODE	COMM_CODE	NAME	SECTOR	SRG	COMM_STRUCTURE	longitude	latitude
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 99% Error 0% Empty < 1%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
	4 distinct, 0 unique	4 distinct, 0 unique	312 distinct, 312 unique	312 distinct, 312 unique	8 distinct, 0 unique	5 distinct, 0 unique	14 distinct, 1 unique	312 distinct, 312 unique	312 distinct, 312 unique
1	Residential	1	BED	BEDDINGTON HEIGHTS	NORTH	ESTABLISHED	1960s/1970s	-114.0850214	51.13163
2	Residential	1	EVN	EVANSTON	NORTH	COMPLETE	2010s	-114.1124526	51.17109
3	Residential	1	KIL	KILLARNEY/GLENGARRY	CENTRE	ESTABLISHED	1950s	-114.1317273	51.03154
4	Residential	1	BRA	BRAESIDE	SOUTH	ESTABLISHED	1960s/1970s	-114.1063659	50.95599
5	Residential	1	BLM	BELMONT	SOUTH	DEVELOPING	BUILDING OUT	-114.0552517	50.86868
6	Industrial	2	FHI	FOOTHILLS	EAST	N/A	EMPLOYMENT	-113.9820389	50.99396
7	Residential	1	RIC	RICHMOND	CENTRE	ESTABLISHED	1950s	-114.1184223	51.02971
8	Residential	1	CHN	CHINATOWN	CENTRE	ESTABLISHED	CENTRE CITY	-114.0626611	51.05088
9	Residential	1	CPF	COPPERFIELD	SOUTHEAST	DEVELOPING	BUILDING OUT	-113.9294735	50.91781
10	Industrial	2	EFV	EAST FAIRVIEW INDUSTRIAL	SOUTH	N/A	EMPLOYMENT	-114.0372414	50.9848
11	Residual Sub Area	4	13H	13H	SOUTH	N/A	UNDEVELOPED	-114.1128018	50.89224
12	Industrial	2	PEG	PEGASUS	NORTHEAST	N/A	EMPLOYMENT	-114.0064889	51.0944
13	Residential	1	ROY	ROYAL OAK	NORTHWEST	COMPLETE	2010s	-114.2197437	51.14266
14	Residential	1	CUR	CURRIE BARRACKS	WEST	ESTABLISHED	BUILDING OUT	-114.1248249	51.01571
15	Residential	1	HKS	HOTCHKISS	SOUTHEAST	DEVELOPING	BUILDING OUT	-113.8962547	50.91419
16	Residential	1	ASP	ASPEN WOODS	WEST	DEVELOPING	BUILDING OUT	-114.2083988	51.0452
17	Residual Sub Area	4	13N	13N	SOUTH	N/A	OTHER	-114.1437681	50.94484
18	Residential	1	HAY	HAYSBORO	SOUTH	ESTABLISHED	1950s	-114.0833403	50.97221
19	Residential	1	CHR	CHRISTIE PARK	WEST	ESTABLISHED	1980s/1990s	-114.1760167	51.04080
20	Residential	1	ALT	ALTADORE	CENTRE	ESTABLISHED	INNER CITY	-114.1018631	51.01595
21	Residential	1	UMR	UPPER MOUNT ROYAL	CENTRE	ESTABLISHED	INNER CITY	-114.0850185	51.02997
22	Residential	1	RED	RED CARPET	EAST	ESTABLISHED	1960s/1970s	-113.9427024	51.04019

3. We renamed the columns appropriately.

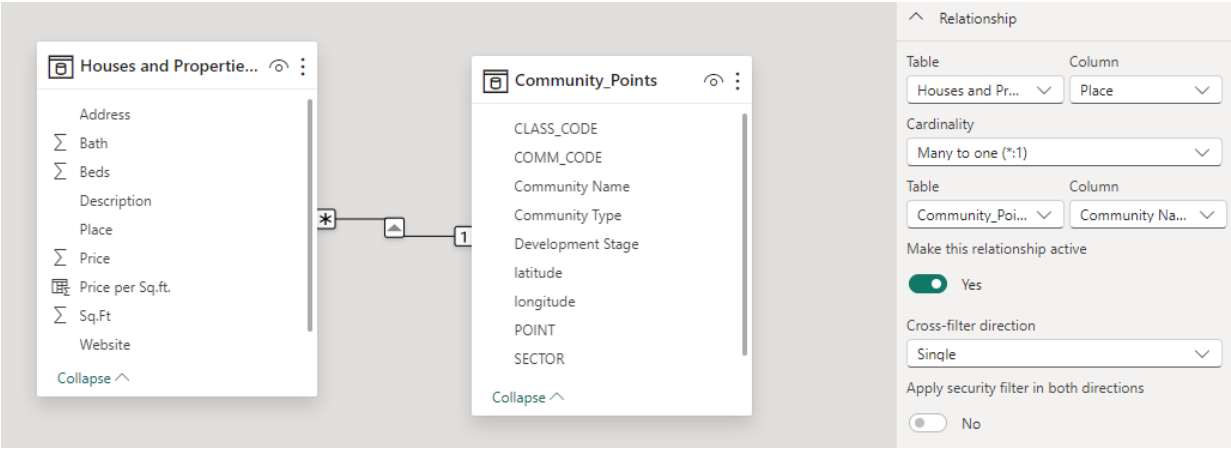


	Community Type	CLASS_CODE	COMM_CODE	Community Name	SECTOR	SRG	Development Stage	longitude	latitude
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 99% Error 0% Empty < 1%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
	4 distinct, 0 unique	4 distinct, 0 unique	312 distinct, 312 unique	312 distinct, 312 unique	8 distinct, 0 unique	5 distinct, 0 unique	14 distinct, 1 unique	312 distinct, 312 unique	312 distinct, 312 unique
1	Residential	1	BED	BEDDINGTON HEIGHTS	NORTH	ESTABLISHED	1960s/1970s	-114.0850214	51.13163
2	Residential	1	EVN	EVANSTON	NORTH	COMPLETE	2010s	-114.1124526	51.17109
3	Residential	1	KIL	KILLARNEY/GLENGARRY	CENTRE	ESTABLISHED	1950s	-114.1317273	51.03154
4	Residential	1	BRA	BRAESIDE	SOUTH	ESTABLISHED	1960s/1970s	-114.1063659	50.95599
5	Residential	1	BLM	BELMONT	SOUTH	DEVELOPING	BUILDING OUT	-114.0552517	50.86868
6	Industrial	2	FHI	FOOTHILLS	EAST	N/A	EMPLOYMENT	-113.9820389	50.99396
7	Residential	1	RIC	RICHMOND	CENTRE	ESTABLISHED	1950s	-114.1184223	51.02971
8	Residential	1	CHN	CHINATOWN	CENTRE	ESTABLISHED	CENTRE CITY	-114.0626611	51.05088
9	Residential	1	CPF	COPPERFIELD	SOUTHEAST	DEVELOPING	BUILDING OUT	-113.9294735	50.91781
10	Industrial	2	EFV	EAST FAIRVIEW INDUSTRIAL	SOUTH	N/A	EMPLOYMENT	-114.0372414	50.9848
11	Residual Sub Area	4	13H	13H	SOUTH	N/A	UNDEVELOPED	-114.1128018	50.89224
12	Industrial	2	PEG	PEGASUS	NORTHEAST	N/A	EMPLOYMENT	-114.0064889	51.0944
13	Residential	1	ROY	ROYAL OAK	NORTHWEST	COMPLETE	2010s	-114.2197437	51.14266
14	Residential	1	CUR	CURRIE BARRACKS	WEST	ESTABLISHED	BUILDING OUT	-114.1248249	51.01571

PROPERTIES
Name: Community_Points
All Properties

APPLIED STEPS
Source
Promoted Headers
Changed Type
Renamed Columns

4. We created Relationships between two data sets using Place and Community Name.



We connected the "Place" column from the "Housing Price & Real Estate - 2023" dataset to the "Community Name" column in the "Community Points" dataset. This link creates a many-to-one relationship, suggesting that several properties listed in the "Housing Price & Real Estate - 2023" dataset could match one community in the "Community Points" dataset.

10. Evaluation:

After reviewing the entire process once more, let's now outline the next steps. With all the necessary data in hand, we are ready to formulate business questions. Subsequently, we'll develop visualizations to address these questions and deploy them in Power BI. Finally, we'll present our analysis through a PowerPoint presentation.

11. Business Questions

1. What is the average price per square foot for homes in different communities, and how does it vary across sectors? And how does the community type (Industrial and Residential) influence the price of homes for sale?

- **Average Price per Square Foot for Homes in Different Communities and Sectors**:
The below dashboard includes a table and a bar chart that display the average price per square foot for various communities across different sectors (CENTRE, EAST, NORTH, NORTHEAST, NORTHWEST, SOUTH, SOUTHEAST, WEST)

- **Variation Across Sectors**:
The bar chart visually represents these differences, showing a wide range of average prices per square foot within sectors.

- **Variation Across Community Type**:

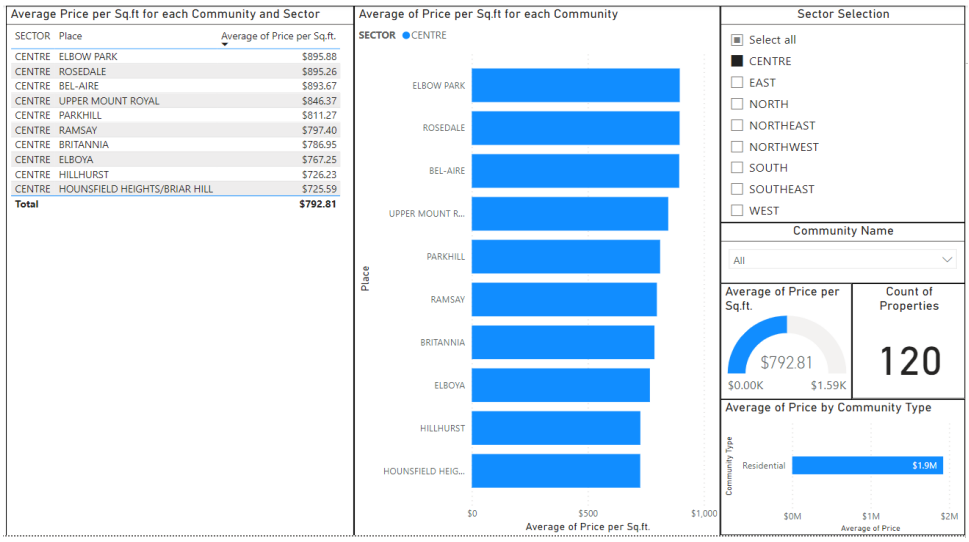
It shows if it's a residential community or industrial based on Average of Price.

- **Count of Properties**:

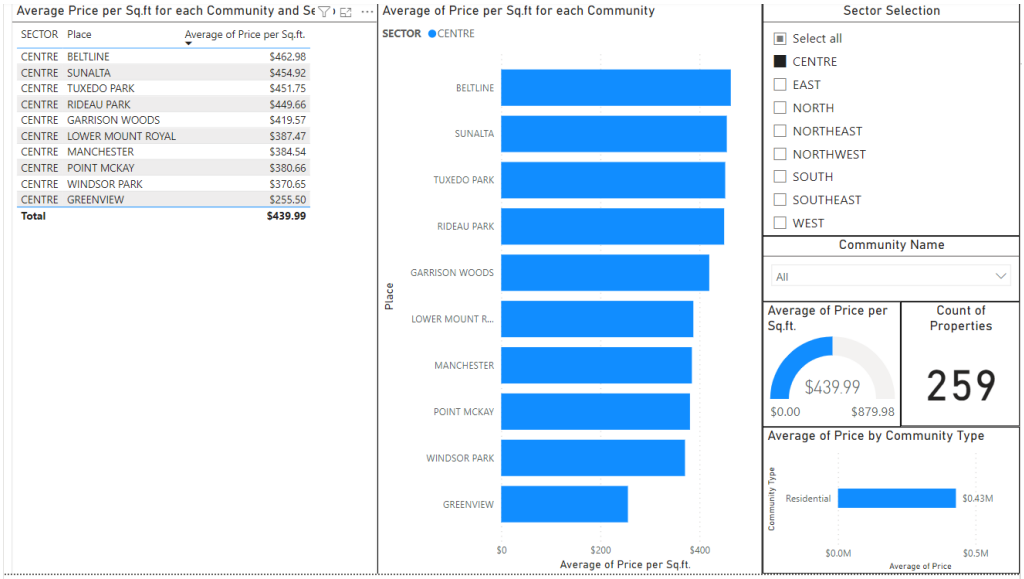
It shows the number of properties listed in that community.

By this dashboard will analyze and answer the above business question

Top 10 and bottom 10 in Centre sector.

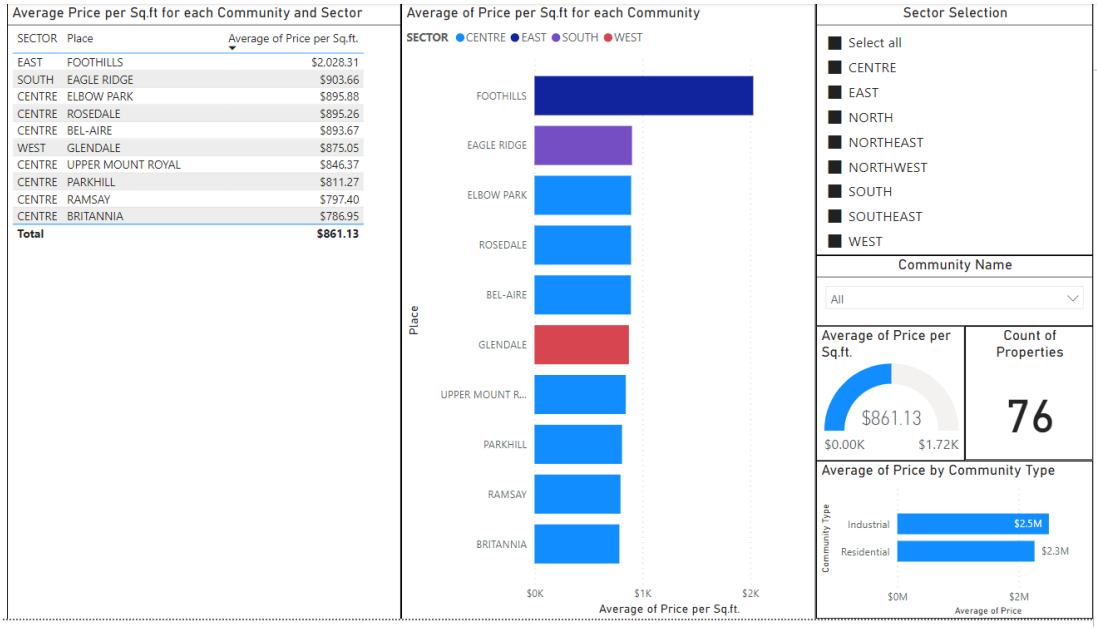


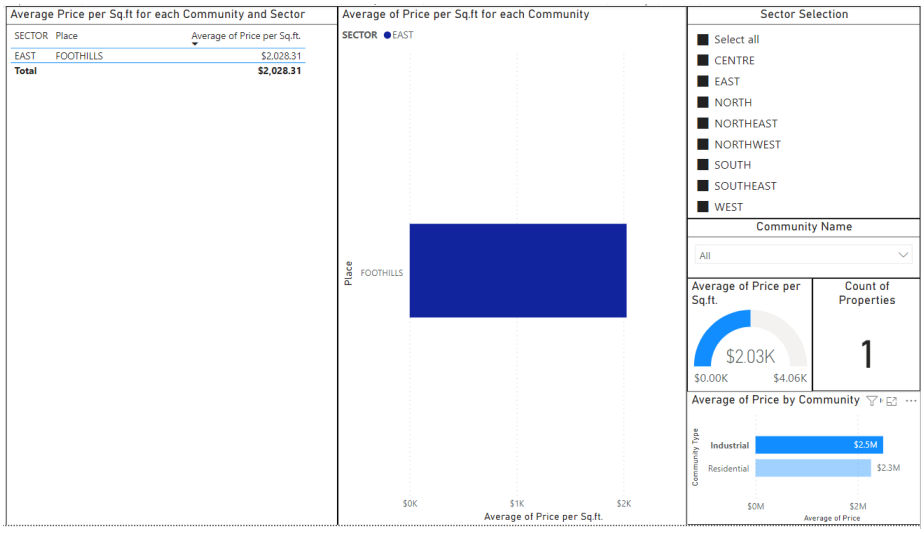
In the above dashboard image we can see the top 10 in the CENTRE sector, Elbow Park has an average price of $895.88 per square foot which is one of the highest in that sector and the average price per sq ft in that Sector for that top 10 is around $792.81 which is residential class and total number of properties listed is 120 and the average of property price is $1.9M.



In the above dashboard image, we can see the bottom 10 in the CENTRE sector, Greenview has an average price of $255.50 per square foot which is one of the lowest in that sector and the average price per sq ft in that sector for that bottom 10 is around $439.99 which is residential class and total number of properties listed is 259 and the average of property price is $0.43M.

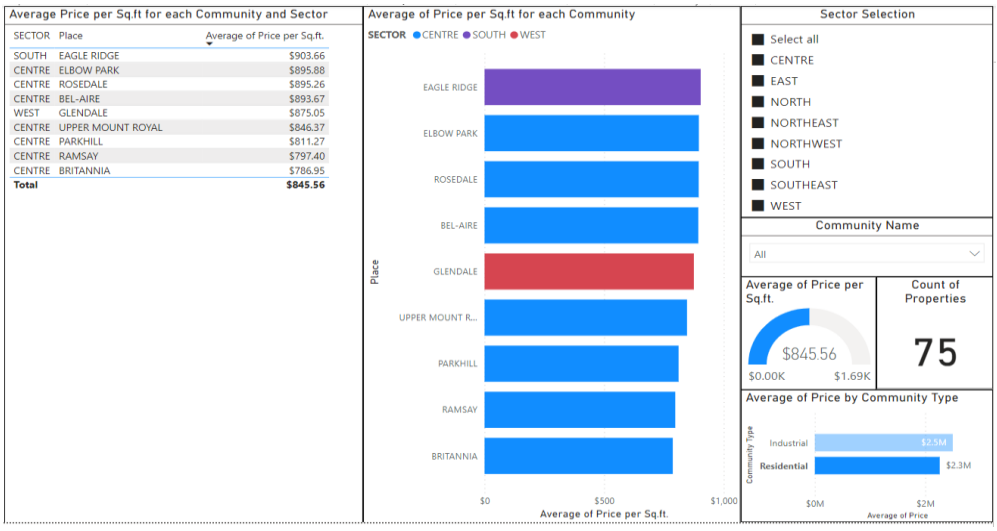
For Top 10 across all sectors





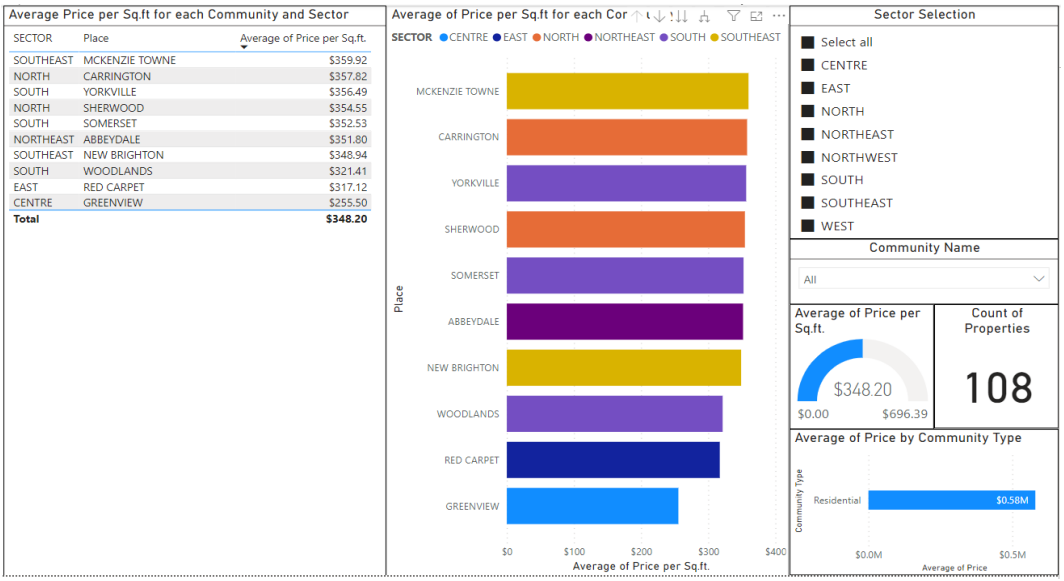
In the above dashboard images if we consider all the sectors instead of just center, we can see the top 10 across all sectors, Foothills Community has an average price of $2028.31 per square foot which is one of the highest across all sectors which is in East sector, and it is an industrial class that's the reason for high price. the average price per sq ft for that top 10 is around $861.13 which comes under both residential and industrial class and total number of properties listed is 10 and the average of property price is $2.3M if it's a residential class and $2.5M in industrial class. Also, we can see that the north, northeast, northwest and southeast are not available in the top 10.

For Top 10 across all sectors in Residential Community type



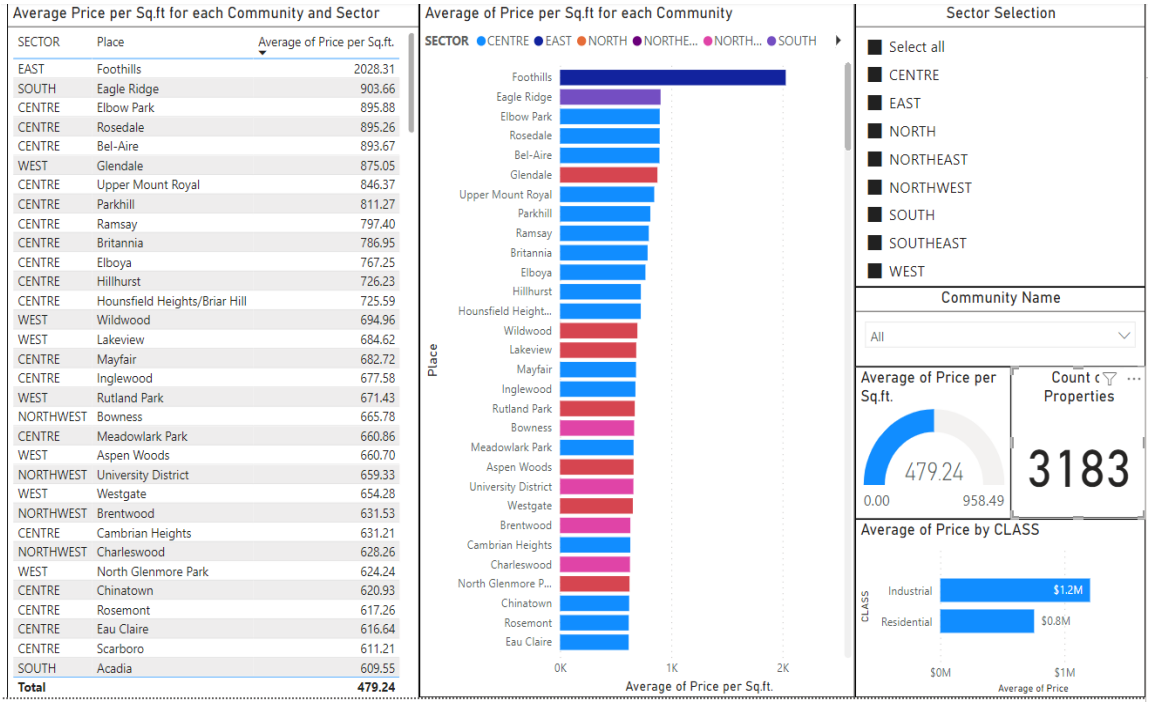
In the above dashboard images if we consider all the sectors instead of just center and ignore the industrial class which is foothill community and only consider residential class, we can see the top 10 across all sectors for residential class, Eagle Ridge has an average price of $903.66 per square foot which is one of the highest across all sector in residential class which is in South sector and the average price per sq ft for that top 10 is around $845.56 which comes under residential and the average of property price is $2.3M.

For Bottom 10 across all sectors



In the above dashboard image if we consider all the sectors instead of just center, we can see the bottom 10 across all sectors, Again Greenview has an average price of $255.50 per square foot which is one of the lowest across all sector even though it's in Centre sector and the average price per sq ft for that bottom 10 is around $348.20 which is residential class and total number of properties listed is 10 and the average of property price is $0.58M. Also, we can see that the west and northwest sector is not available in the bottom 10.

We can see by the below dashboard without any filters:



Industrial properties have a higher average price at $1.2 million, whereas residential properties have a lower average price at $0.8 million. This suggests that industrial properties tend to be priced higher per square foot compared to residential properties.

Overall, the dashboard provides a comprehensive view of how home prices per square foot vary across different communities and sectors, and how community type affects these prices. By using this buyer can decide where they want to purchase the kind of property they are looking for and the average price per square foot and total properties available to look at in that community, etc.

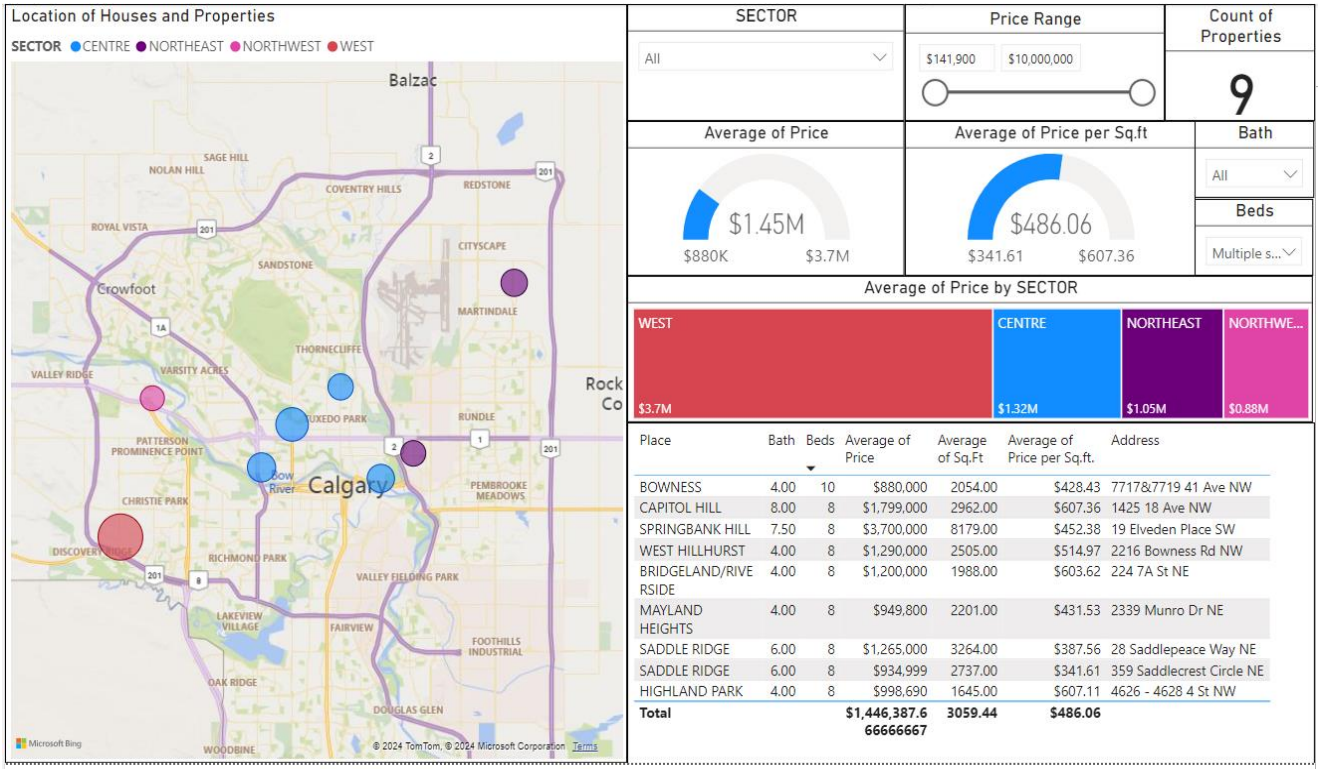
This analysis can be crucial for strategic decisions related to investment or development in specific sectors.

2. **In pursuit of investment opportunities, what type of property would be optimal in terms of maximizing the number of bedrooms? Additionally, in which community is such a property typically located, and what is the associated selling price?**

   If we look at properties with the maximum number of bedrooms listed, with multiple selection from 10 and 8 bedrooms with all number of bathrooms with no price range limit.

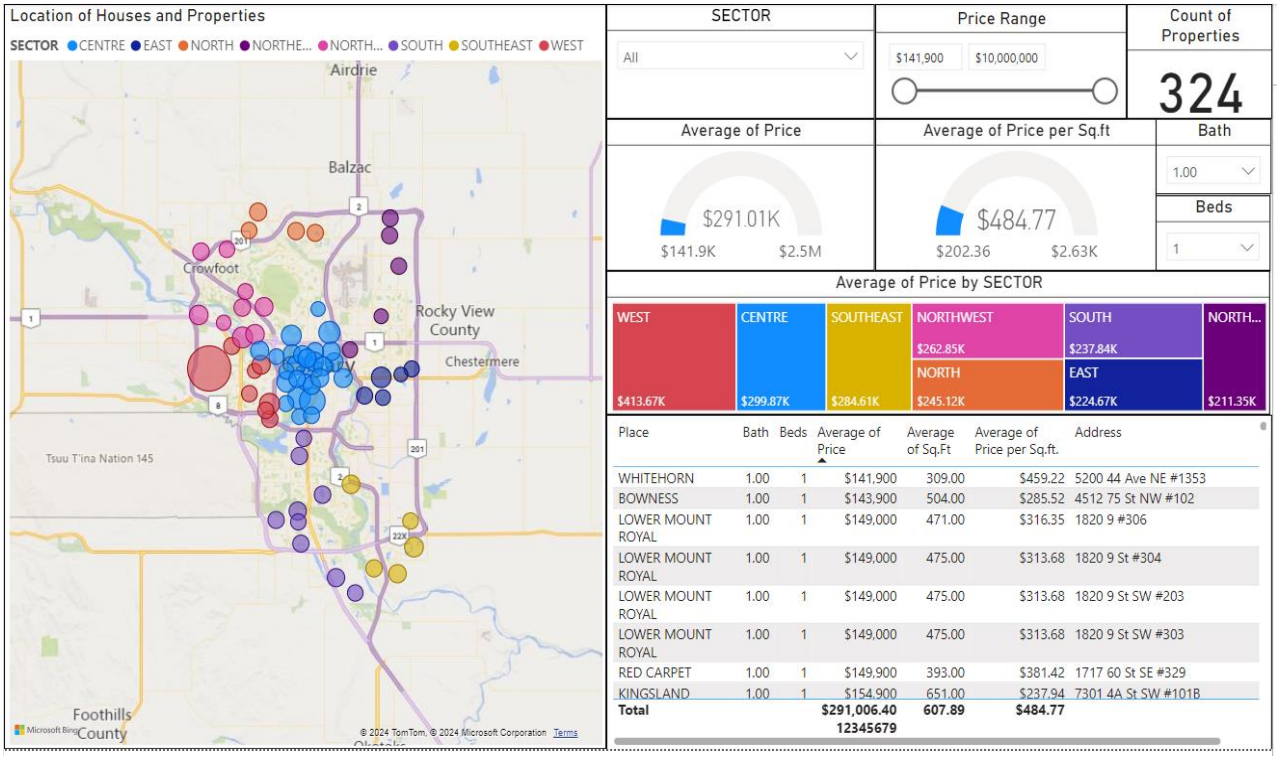
   We can see from the below dashboard that there are 9 such properties which are suitable for investment with max of bedrooms in Centre, Northeast, Northwest, and west sectors.

   Using this dashboard, stake holders can compare it with other properties which are similar and invest in these properties and list it for rent so that they can get good rental income.



Place	Bath	Beds	Average of Price	Average of Sq.Ft	Average of Price per Sq.ft.	Address
BOWNESS	4.00	10	$880,000	2054.00	$428.43	7717&7719 41 Ave NW
CAPITOL HILL	8.00	8	$1,799,000	2962.00	$607.36	1425 18 Ave NW
SPRINGBANK HILL	7.50	8	$3,700,000	8179.00	$452.38	19 Elveden Place SW
WEST HILLHURST	4.00	8	$1,290,000	2505.00	$514.97	2216 Bowness Rd NW
BRIDGELAND/RIVERSIDE	4.00	8	$1,200,000	1988.00	$603.62	224 7A St NE
MAYLAND HEIGHTS	4.00	8	$949,800	2201.00	$431.53	2339 Munro Dr NE
SADDLE RIDGE	6.00	8	$1,265,000	3264.00	$387.56	28 Saddlepeace Way NE
SADDLE RIDGE	6.00	8	$934,999	2737.00	$341.61	359 Saddlecrest Circle NE
HIGHLAND PARK	4.00	8	$998,690	1645.00	$607.11	4626 - 4628 4 St NW
Total			**$1,446,387.66666667**	**3059.44**	**$486.06**	

3. For a single individual with the financial means to purchase a property featuring one bedroom and one bathroom, what would be the anticipated cost of such a unit?



Place	Bath	Beds	Average of Price	Average of Sq.Ft	Average of Price per Sq.ft.	Address
WHITEHORN	1.00	1	$141,900	309.00	$459.22	5200 44 Ave NE #1353
BOWNESS	1.00	1	$143,900	504.00	$285.52	4512 75 St NW #102
LOWER MOUNT ROYAL	1.00	1	$149,000	471.00	$316.35	1820 9 #306
LOWER MOUNT ROYAL	1.00	1	$149,000	475.00	$313.68	1820 9 St #304
LOWER MOUNT ROYAL	1.00	1	$149,000	475.00	$313.68	1820 9 St SW #203
LOWER MOUNT ROYAL	1.00	1	$149,000	475.00	$313.68	1820 9 St SW #303
RED CARPET	1.00	1	$149,900	393.00	$381.42	1717 60 St SE #329
KINGSLAND	1.00	1	$154,900	651.00	$237.94	7301 4A St SW #101B
Total			**$291,006.40**	**607.89**	**$484.77**	
			12345679			

Based on the above dashboard we can see that a single individual who is looking for a 1 bed 1 bath property they have total of 324 properties listed based on the price range 141,900 $ to 2,50,0000 $ and 307 sq. ft. to 1057 sq. ft. to choose from.

With this information, stakeholders can make selections based on sq. ft., community, and price to determine the most suitable option.

12. Deployment:

As Mentioned in Evaluation we deployed all the Power Bi visuals in Power Bi Server and Presentation will and other related Files are uploaded in Brightspace which can be accessed by the stakeholders to analyze and decide on their investment.
Link for Power Bi Server : Final assignment - NW - Power BI

https://app.powerbi.com/groups/me/reports/84bd6331-498c-42da-9409-870add279e4a/ReportSection?experience=power-bi

13. Conclusion:

Through our comprehensive analysis of the Canadian housing market in 2023, utilizing datasets from Kaggle and Calgary Open Data, we have gained valuable insights into housing price trends, regional variations, and the impact of community types on property values. We employed the CRISP-DM methodology, ensuring a structured approach from data understanding to visualization and analysis.

Key Findings:

1. **Regional and Sectoral Variations:** Our analysis revealed significant disparities in housing prices across different regions and sectors. From affluent neighborhoods commanding high prices per square foot to more affordable areas, buyers and investors have a wide range of options to consider.
2. **Influence of Community Type:** We observed that community type, whether residential or industrial, plays a crucial role in determining property values. Industrial properties tended to have higher average prices, highlighting the importance of understanding the specific dynamics of each community.
3. **Optimal Property Types for Investment:** Our exploration of properties with the maximum number of bedrooms indicated potential investment opportunities in various sectors. Understanding the demand for specific property types can guide investors in making informed decisions.
4. **Affordability and Accessibility:** For individuals seeking properties with one bedroom and one bathroom, our analysis identified a range of options across different price points and sizes. This information can empower prospective buyers to navigate the market efficiently.

13. Recommendations:

1. **Diversification of Investment Portfolio:** Investors should consider diversifying their portfolios across different regions and property types to mitigate risks and capitalize on emerging opportunities. Understanding market dynamics and demographic trends is essential for strategic decision-making.
2. **Community-Level Analysis:** Further research into community-level factors, such as infrastructure development, amenities, and future growth projections, can provide deeper insights into long-term investment prospects. Collaborating with local stakeholders and real estate professionals can enhance understanding and decision-making.
3. **Continuous Monitoring and Adaptation:** The real estate market is dynamic, influenced by economic shifts, regulatory changes, and societal trends. Regular monitoring of market indicators and updating analytical models will ensure relevance and accuracy in decision support.
4. **Investment in Data Analytics:** Given the importance of data-driven insights in the real estate sector, investing in advanced analytics tools and expertise can provide a competitive edge. Leveraging predictive modeling and machine learning algorithms can enhance forecasting capabilities and risk assessment.
5. **Improvement to Dataset:** we suggest creating a dataset based on postal codes with sector data for Calgary along with nearby hotspots or malls. This will provide more accurate and localized insights into housing market trends. Collaborating with local authorities can ensure data accuracy. Integrating this dataset will help stakeholders make informed decisions tailored to specific neighborhoods and sectors within Calgary, enhancing investment strategies.

---Thank You---