RENTAL PRICE PREDICTION

Leveraging machine learning to analyze Vancouver’s rental market.

Hi, I'm Pam!

- Business Engineering - > Data Science Co-Op
- 5 year - Experience in Digital Marketing & Analysis for Startups, SaaS, Service, Retail Industries.
- Tech Stack: Python, SQL, PostgreSQL, BigQuery
- Project: Python for ML, Looker

WHY?

RENTING A PLACE IN VANCOUVER IS A CHALLENGE

- Landlords are **increasing rents by 23.5%** when apartments become vacant.
- Vancouver's rental market has a **0.8% vacancy rate** in purpose-built rentals.
- CMHC's Fall 2024 Rental Market Report shows that **rising rents and low vacancy rates** discourage tenant mobility in Vancouver.

1 - CBC **|** https://www.cbc.ca/news/canada/british-columbia/rent-hike-23-per-cent-1.7295152
2 - CMHC **|** 2024-02-12-memo-updated-rental-market-data-from-cmhc-for-2024
3 - Fall 2024 Rental Market Report **|** CMHC - Fall 2024 Rental Market Report | CMHC

Where are you looking for rentals?

GOAL

Predict rent prices in Vancouver using Machine Learning.

HOW?

1. Web Scraping

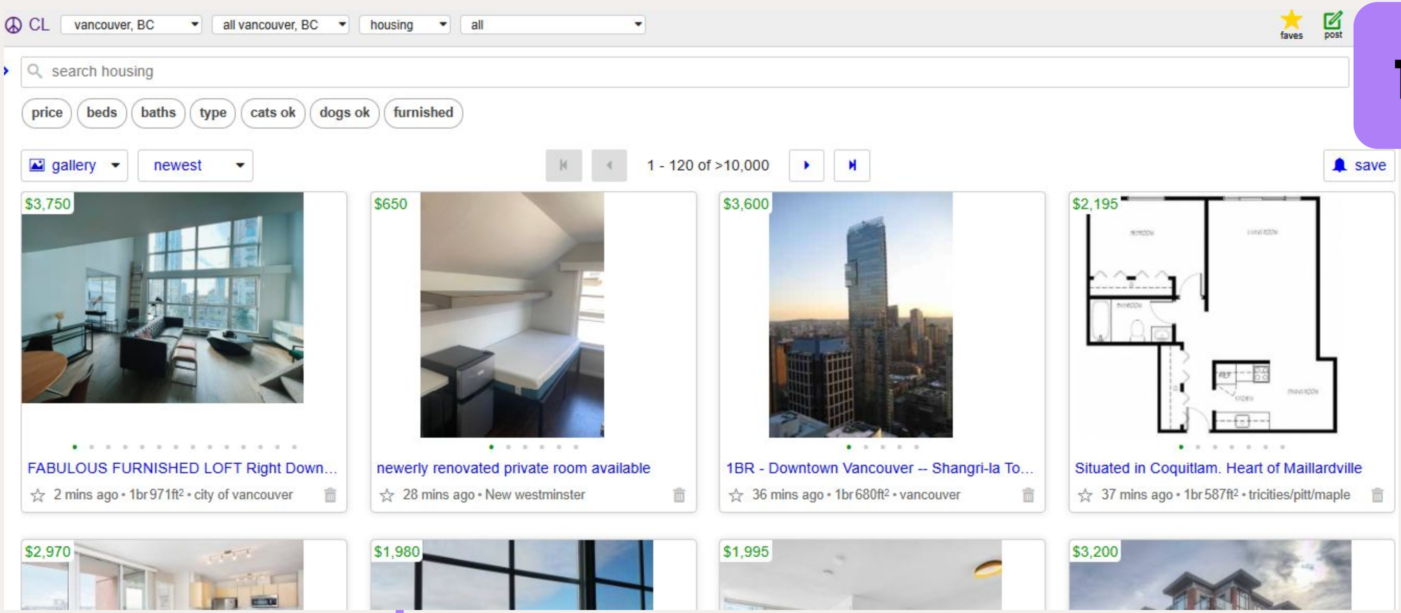
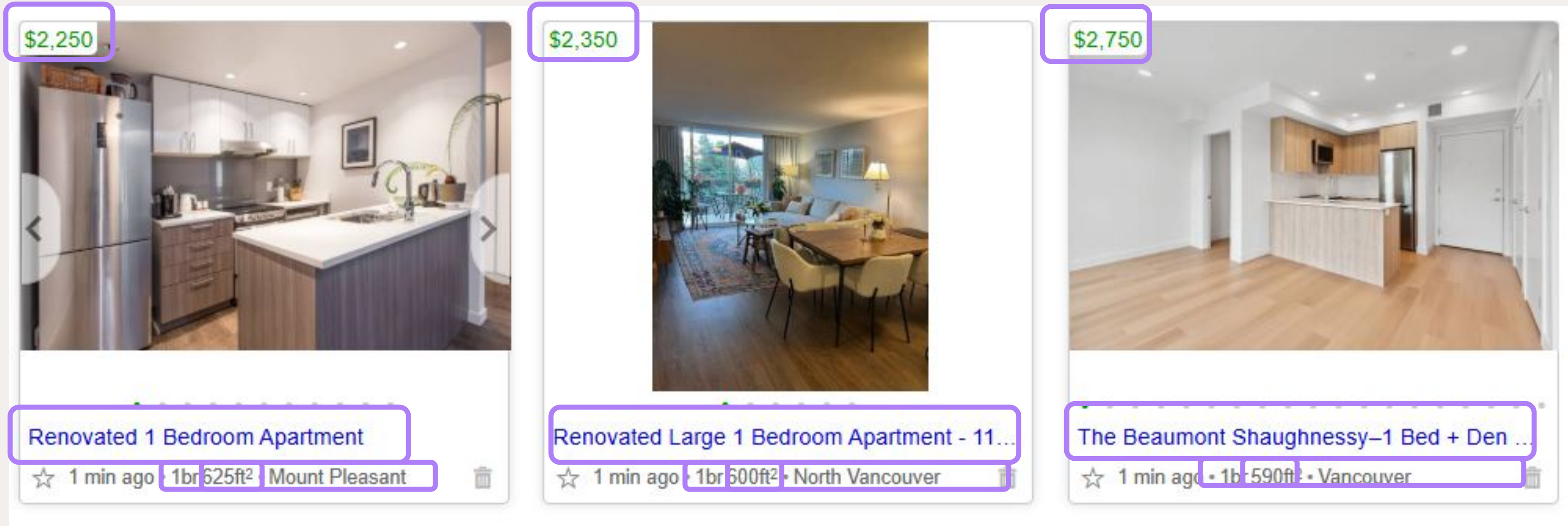


Image	Title_URL	Label	meta	postbedrooms	postsqft	Price
	https://vancouver.craigslist.org/rds/apa/d/white-rock-2b-1b-secure-building/7818386784.html	2B/1B, Secure Building, Situated in White Rock!	5 mins ago2br919ft21371 Blackwood Street, White Rock, BC	2br	919ft2	$2,695
https://images.craigslist.org/d/7821650955/00O0O_1dNwbSIMEiP_0CI0t2_600x450.jpg	https://vancouver.craigslist.org/van/apa/d/vancouver-1bedroom-available-in-cozy/7821650965.html	1Bedroom available in Cozy Laneway house 2 balcony, 1 den, 1 bath	7 mins ago1brVancouver	1br		$1,195
https://images.craigslist.org/d/7815177258/01010_gE8D6R4PdZP_0CI0pO_600x450.jpg	https://vancouver.craigslist.org/rds/apa/d/port-coquitlam-house-rent/7815177258.html	HOUSE RENT	9 mins ago4br1250ft2Port Coquitlam	4br	1250ft2	$3,600
https://images.craig						

Data


$2,250
Renovated 1 Bedroom Apartment
1 min ago 1br 625ft2 Mount Pleasant
$2,350
Renovated Large 1 Bedroom Apartment - 11...
1 min ago 1br 600ft2 • North Vancouver
$2,750
The Beaumont Shaughnessy–1 Bed + Den ...
1 min ago • 1br 590ft2 • Vancouver

2. Data Cleanup

```python
# Lower case the "location" column to clean it.
df['location'] = df['location'].str.lower()

# Replace al strings that have a "Vancouver" with just "Vancouver"
df['location'] = df['location'].str.replace(r'.*surrey.*', 'surrey', regex=True)
df['location'] = df['location'].str.replace(r'.*\b(burnaby|brentwood|metrotown)\b.*', 'burnaby', regex=True)
df['location'] = df['location'].str.replace(r'.*richmond.*', 'richmond', regex=True)
df['location'] = df['location'].str.replace(r'.*delta.*', 'delta', regex=True)
df['location'] = df['location'].str.replace(r'.*maple ridge.*', 'maple ridge', regex=True)
df['location'] = df['location'].str.replace(r'.*pitt meadows.*', 'pitt meadows', regex=True)
df['location'] = df['location'].str.replace(r'.*white rock.*', 'white rock', regex=True)
df['location'] = df['location'].str.replace(r'.*langley.*', 'langley', regex=True)
df['location'] = df['location'].str.replace(r'.*coquitlam.*', 'coquitlam', regex=True)
df['location'] = df['location'].str.replace(r'.*tsawwassen.*', 'tsawwassen', regex=True)
df['location'] = df['location'].str.replace(r'.*port moody.*', 'port moody', regex=True)
df['location'] = df['location'].str.replace(r'.*new westminster.*', 'port moody', regex=True)
df['location'] = df['location'].str.replace(
    r'.*\b(north|west) vancouver\b.*', r'\1 vancouver', regex=True  # Preserve special cases
)

# Replace remaining patterns containing "Vancouver" with "Vancouver"
df['location'] = df['location'].str.replace(
    r'.*vancouver.*', 'vancouver', regex=True)

# If the value is not then turn into vancouver
df['location'] = df['location'].apply(lambda x: x if x in [
    'surrey', 'burnaby', 'richmond', 'delta', 'maple ridge', 'pitt meadows', 'white rock', 'langley', 'coquitlam'
    , 'tsawwassen', 'port moody', 'vancouver'] else 'vancouver')

df['location'].unique()
```

```
Image           object
Title_URL       object
Label           object
meta            object
postbedrooms    object
postsqft        object
Price           object
zone_name       object
dtype: object
```

```
Image           object
Title_URL       object
Label           object
meta            object
postbedrooms   float64
postsqft       float64
Price          float64
location        object
dtype: object
```

2. Data Cleanup

```
# find empty values
df.isnull().sum()
```

```
Image          333
Title_URL        0
Label            0
meta             0
postbedrooms   391
postsqft       971
Price            1
dtype: int64
```

The null values appear because the user left these sections of the ad on Craigslist unfilled.

EDA

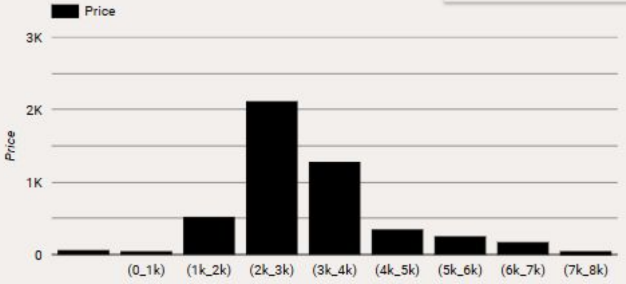
Exploratory Data Analysis

RENT ANALYSIS VANCOUVER

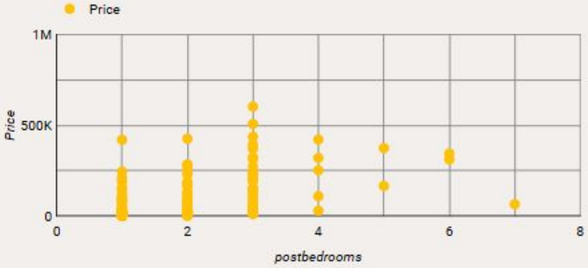
Record Count 4.8K	postsqft 1,057.87
Price 3.3K	postbedrooms 2.06

Price Distribution Histogram

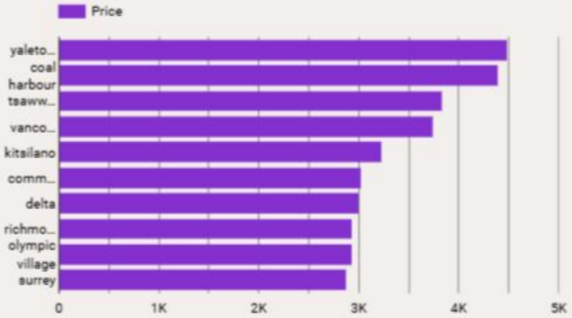
Price Bin



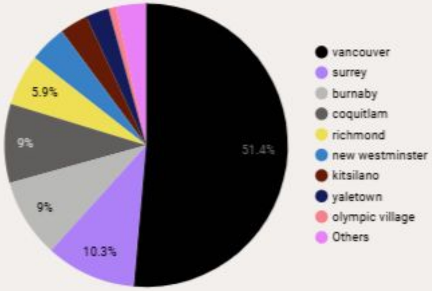
Price vs. Square Footage Scatter Plot



Price by Location Bar Chart:



Listings by Location Pie Chart



Let’s go to:

Looker Studio

4. Encoding

Label Encoding

```python
# Initialize the label encoder
label_encoder = LabelEncoder()

# Apply label encoding to the 'zone_name' column
df_ml['location_encoded'] = label_encoder.fit_transform(df_ml['location'])
df_ml = df_ml.drop(columns=['location'])
df_ml.head(2)
```

	postbedrooms	postsqft	Price	location_encoded
2	4.0	1250.0	3600.0	1
3	3.0	1250.0	3600.0	1

Vancouver is priority.

Test_size = 0.2 means **80%** of the data will be used for training and **20% for testing.**

5. Feature Scaling

```python
# Initialize StandardScaler
scaler = StandardScaler()

scaler = StandardScaler()
X_train[["postbedrooms", "postsqft"]] = scaler.fit_transform(X_train[["postbedrooms", "postsqft"]])
X_test[["postbedrooms", "postsqft"]] = scaler.transform(X_test[["postbedrooms", "postsqft"]])
```

Scaling ensures all the features contribute equally to the model.

ML
MODEL

6. Machine Learning

RandomForestRegressor model

```python
# Train a Random Forest Regressor
rf_model = RandomForestRegressor(random_state=42, n_estimators=100)
rf_model.fit(X_train, y_train)
y_pred_rf = rf_model.predict(X_test)

# Evaluate the model
mse_rf = mean_squared_error(y_test, y_pred_rf)
r2_rf = r2_score(y_test, y_pred_rf)

print(f'Mean Squared Error (Random Forest): {mse_rf}')
print(f'R-squared (Random Forest): {r2_rf}')
```

```
Mean Squared Error (Random Forest): 998824.9979292797
R-squared (Random Forest): 0.7226820599458545
```

MSE = 998824

R^2 = 0.72

RESULTS

```python
# Create new data for prediction
new_data = pd.DataFrame({
    "postbedrooms": [2, 2, 2],
    "postsqft": [800, 800, 800],
    "location": ["vancouver", "burnaby", "coquitlam"]
})
```

```
evaluation_results
```

```
{'Random Forest MSE': 998824.9979292797,
 'Random Forest R²': 0.7226820599458545,
 'Predicted Rental Prices (Random Forest)': [2751, 2173, 2032]}
```

Predicted Rental Prices (RandomForest)

Vancouver: $ 2751

Burnaby: $ 2173

Coquitlam: $ 2032

CONCLUSIONS

HOT RENTALS

1 - 3 bedroom houses / apartments.

VANCOUVER

Can have the <u>lowest price</u> but also the <u>highest price</u> for the same amount of bedrooms.

OPTIONS

<u>Affordable</u> rents in cities nearby.

TO LOOK FORWARD

2025 New research(CIBC): Growth in Unit compilations well outpaced **population growth.**

- International student permits cap (newcomers)
- 45% less (fall of 2024) University in-person enrollments.

2025 might be a good year to look around for rents.

https://globalnews.ca/news/10738537/universities-canada-international-student-enrolment-drop/

Further Projects:

More data

Effect of new regulations

Github repository

Pamela Mollinedo

pamkmoll@gmail.com

+1 7782513491

THANKS!