

The background image is a wide-angle aerial photograph of the Vancouver skyline during sunset. The city's dense urban area is visible in the foreground and middle ground, with numerous skyscrapers and residential buildings. In the background, the Coast Mountains are visible under a clear sky. The water of the Burrard Inlet and English Bay is in the lower left, with several boats and yachts visible.

# Forecasting Compliance: Building a Regression Model to Predict Airbnb Licensing in Vancouver

CASA0007 | QUANTITATIVE METHODS

BOHAO SU, JUANES LAMILLA, NIKHIL DESAI, SIMON SONG

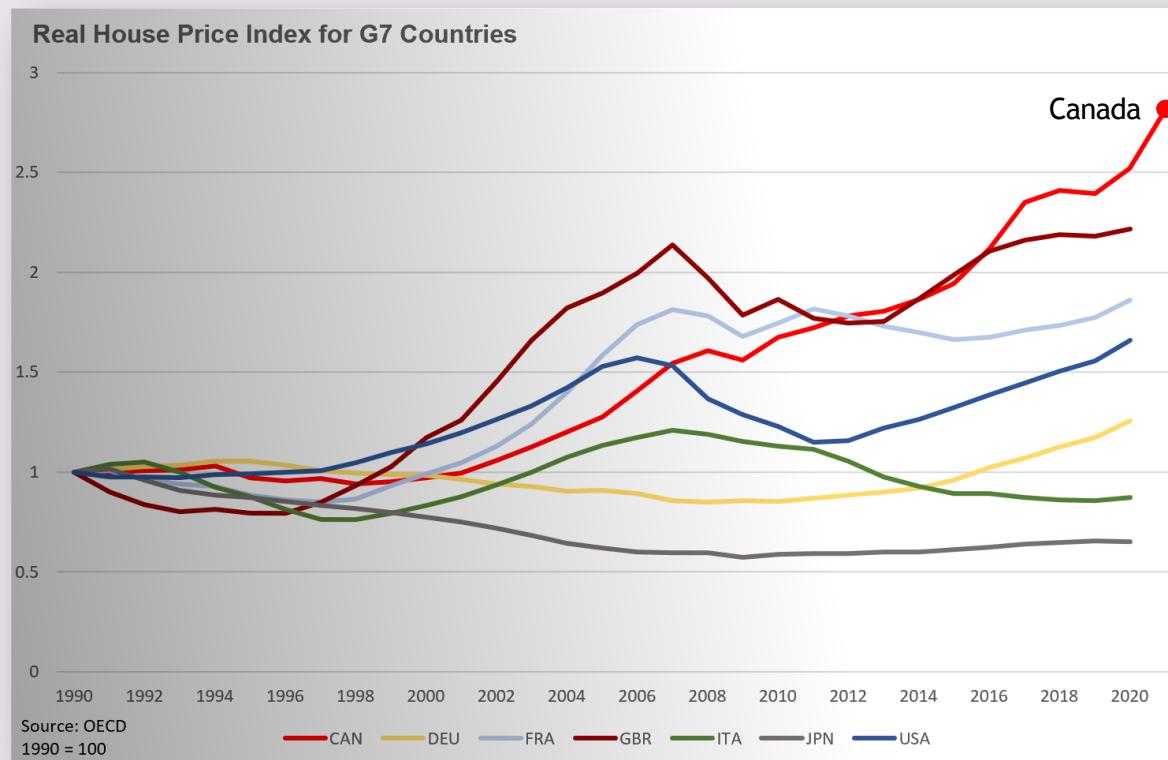


# Agenda

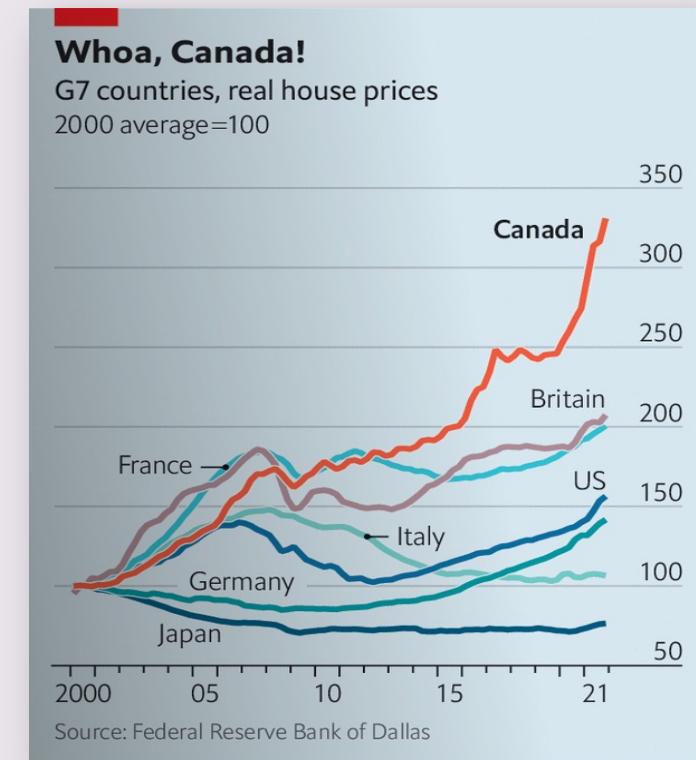
Introduction	<p>An overview of our research topic</p> <p>Why we chose to use <i>Inside Airbnb Data</i></p>	<ul style="list-style-type: none"><li>Background information on YVR and recent <i>Short-Term Rental Accommodations Act</i></li><li>Introduce our dataset and reasons why we chose it</li></ul>
Data Cleaning & Model	<p>An overview of how we have processed the data</p> <p>How we decided which fields to use</p> <p>An overview of the nature of our data</p>	<ul style="list-style-type: none"><li>A brief overview of how we conducted the preprocessing of our dataset</li><li>Overview of our model and how we identified the most relevant field for our analysis</li><li>Small demo of Web app</li><li>Explaining the nature of our data using: assumptions, correlation analysis, correlation matrices and histograms</li></ul>
Conclusion	<p>Summary of our project and policy implications</p>	<ul style="list-style-type: none"><li>Talk about the relevance of our research, what we found and the implications</li></ul>

# Over the past 20 years the Canadian housing market has experienced a rapid price increase

## G7 Countries Housing Price Index and Real House Prices

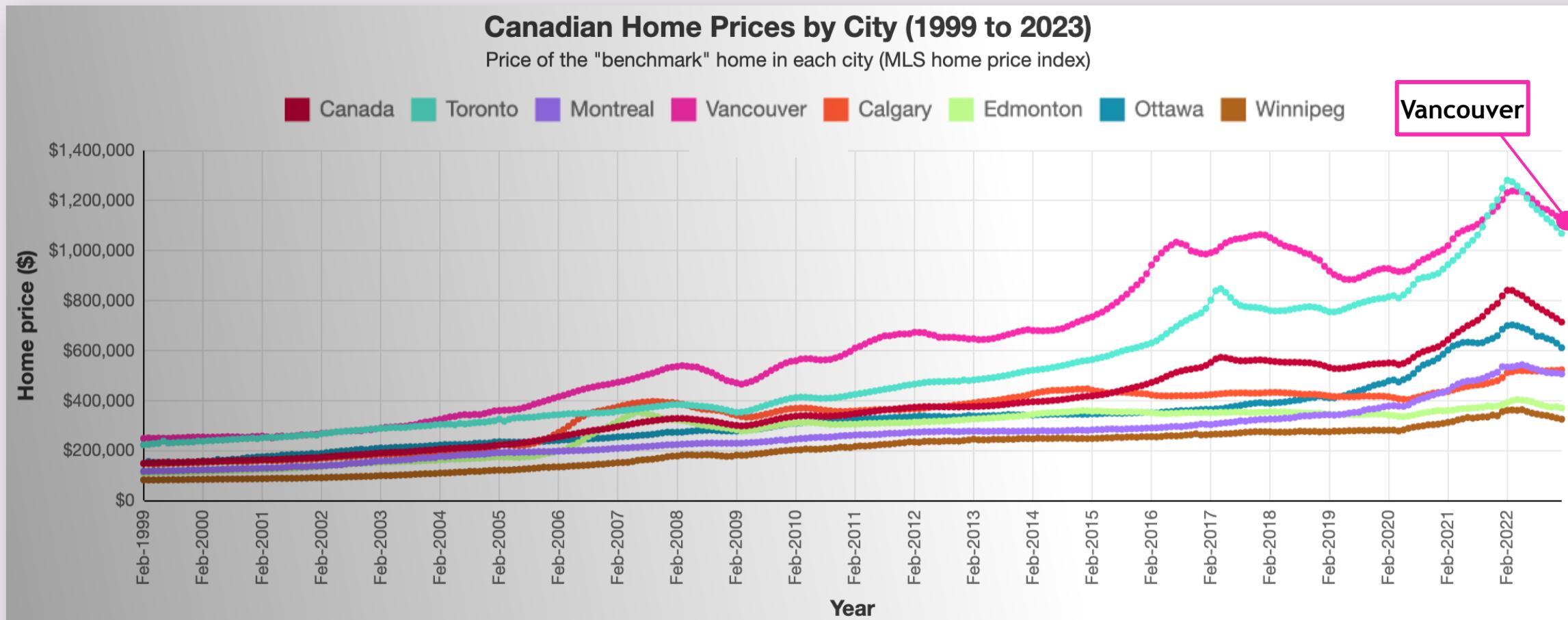


<sup>1</sup> OECD



<sup>2</sup> Federal Reserve Bank of Dallas (2022)

This is experienced most significantly in large cities, like Vancouver. Short term rentals have been blamed



<sup>1</sup> MLS Home Price Index, Seasonally Adjusted Values - Canadian Real Estate Association

# In response, Vancouver has restricted the rights of short-term rental companies culminating in the *Short-Term Rental Accommodations Act* in October 2023



(Ink Drop/Shutterstock)

RENTING

## British Columbia Joins The War On Airbnb, Short-Term Rentals

The Province will introduce the Short-Term Rental Accommodations Act with actions similar to what other governments in North America and beyond have done.

By Howard Chai    October 16, 2023 03:29 pm

### Short-Terms to Long-Terms<sup>1</sup>

Like many other governments in North America, the Province is cracking down on short-term rentals in hopes of converting a large portion — if not the majority — of those short-term rental homes into long-term housing that will aid its housing shortages.

To accomplish this, the Province will now require short-term rentals in municipalities with a population of over 10,000 people — exempting resort communities like Whistler — to be offered only in a host's principal residence, including those with a legal secondary suite. According to [Inside Airbnb](#), 32 hosts in Vancouver, as of September 6, 2023, have more than 10 listings on Airbnb, with nine hosts listing over 30 and one host with 123 listings. The principal residence requirement will address this small set of hosts who operate multiple short-term rentals, with the hope being that all those other listings will become long-term homes.

<sup>1</sup> Storeys: British Columbia Joins The War On Airbnb, Short-Term Rentals, Oct 16, 2023

# 'This is a very serious issue': Growing concerns over illegal Airbnb listings flooding the Vancouver market<sup>1</sup>

City Councillor Lenny Zhou says he's working with staff to figure out just how many illegal short-term rental units are operating in Vancouver.

"Some people are actually not using their principal residence, (they're) using their investment unit as a short-term rental. That's not right," Zhou told CTV News. "This is a very serious issue. For one, our city is experiencing a housing crisis and some people are taking advantage of the system and trying to make some money using some illegal short-term rental."

Under [City of Vancouver regulations](#), you need a valid business licence to operate a short-term rental, and it's illegal to operate a short-term rental that isn't your principal residence.

According to city staff, there are 4,084 active listings for short-term rentals in Vancouver. So far in 2023, 132 licenses have been suspended, 120 violation tickets were issued and 54 units were flagged for investigations and audits.

The current fine is \$1,000 per violation, the maximum allowed under provincial law. Zhou believes that needs to change.

"This is too low," said Zhou. "For short-term rentals you could probably make up the money in 2-3 days."



<sup>1</sup> CTV News Vancouver: *Growing concerns over illegal Airbnb listings flooding the Vancouver Market*. Oct 2023

We have chosen  
to use data from  
*Inside Airbnb*<sup>1</sup>

<sup>1</sup> <http://insideairbnb.com/>



## Vancouver

Filter by:

Vancouver

**6,695**

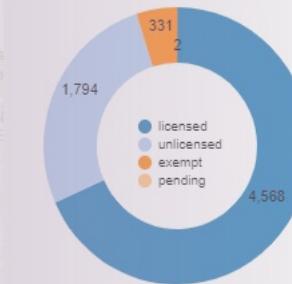
out of 6,695 listings (100.0%)



## Licenses

 Only unlicensed**26.8%**

unlicensed

1,794 (26.8%)  
unlicensed4,568 (68.2%)  
licensed331 (4.9%)  
exempt2 (0.0%)  
pending

## Short-Term Rentals

 Only short-term rentals**64.2%**

short-term rentals

4,300 (64.2%)  
short-term rentals2,395 (35.8%)  
longer-term rentals

The housing policies of cities and towns can be restrictive of short-term rentals, to protect housing for residents.

By looking at the "minimum nights" setting for listings, we can see if the market has shifted to longer-term stays. Was it to avoid regulations, or in response to changes in travel demands?

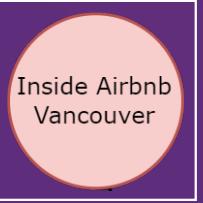
In some cases, Airbnb has moved large numbers



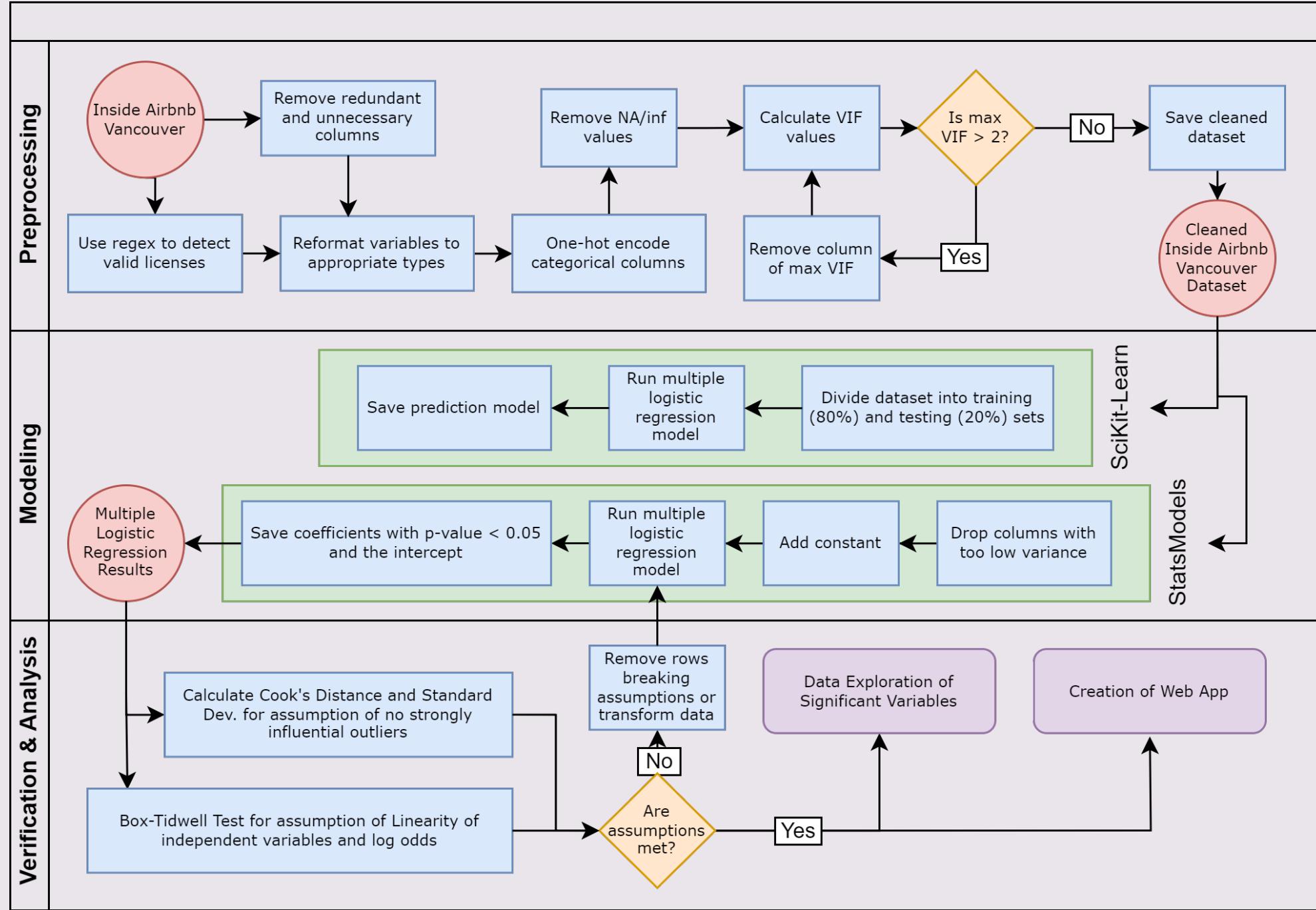
# Predicting Compliance: Building a Multiple Logistic Regression Model to Predict Airbnb Licensing in Vancouver

---

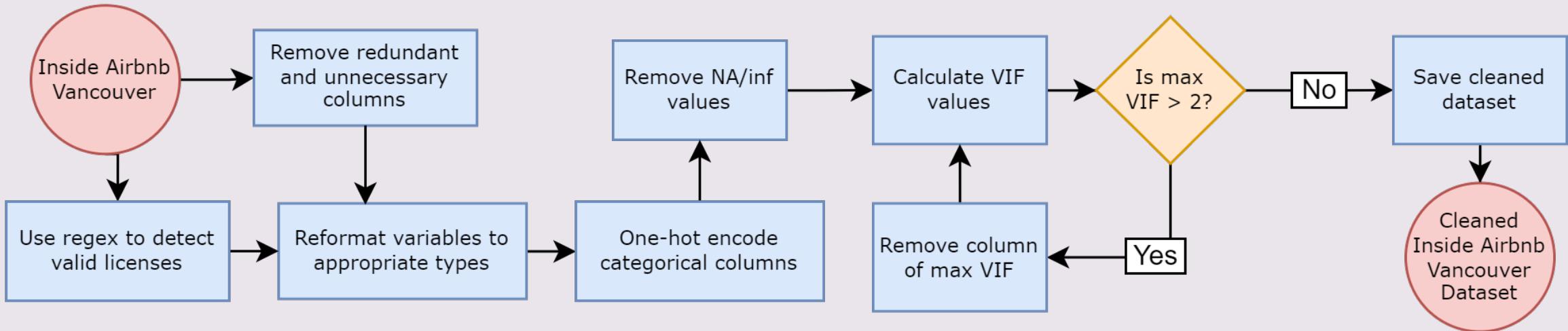
1. Identify fields correlated to licensure using multiple logistic regression
2. Determine how accurate key fields are in predicting whether Vancouver listings are licensed



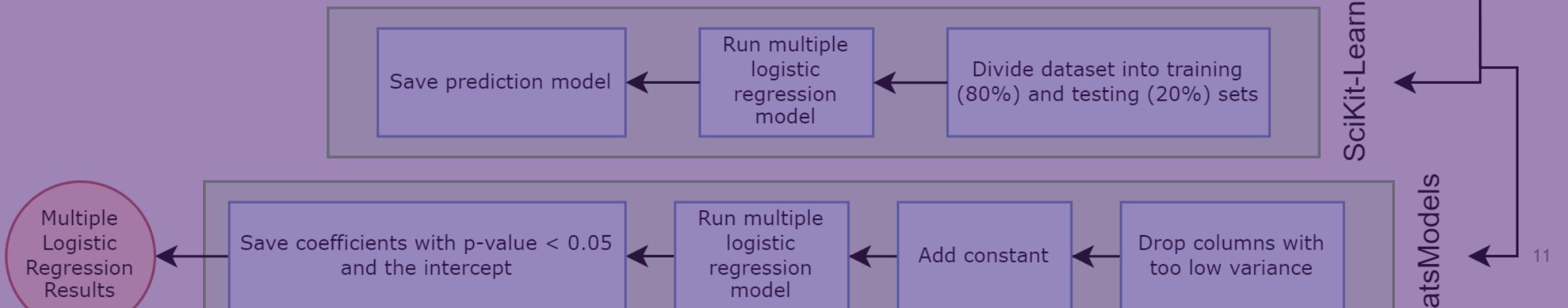
# Data Cleaning and Modeling



## Preprocessing



## Modeling

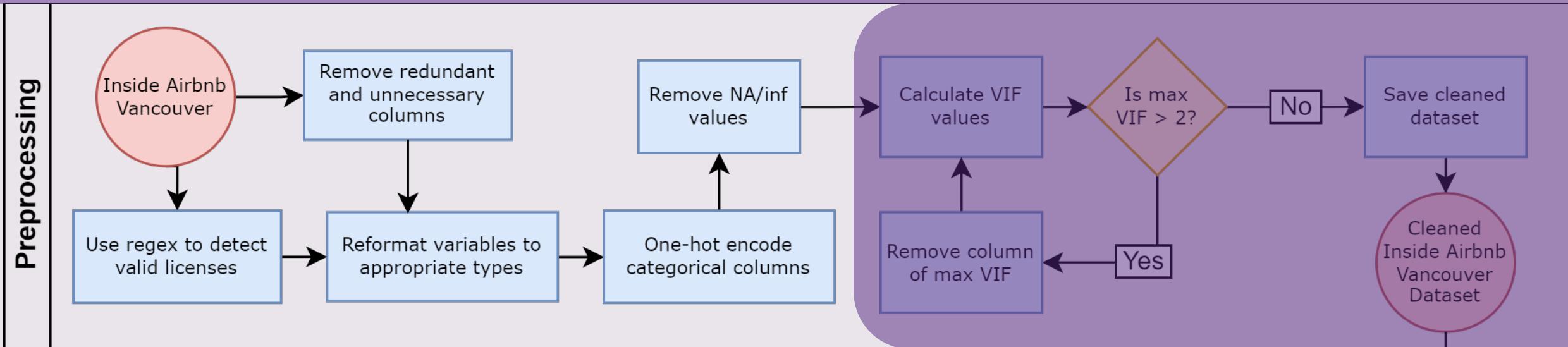


"This apartment rents for one-month blocks of time.  
This is due to new Vancouver city regulations."

### Textual Description

'latitude', 'longitude': 49°17'N, 123°07'W  
'host\_id': 70437, 63238  
'host\_url' : www.airbnb.com/users/show/51466

### Apparently Non-related



'Price': \$150.00, \$350.00

**Continuous**  
(Float & Int)

**Boolean Type**

**'Instant\_Bookable'**: "t/f" → "1/0"

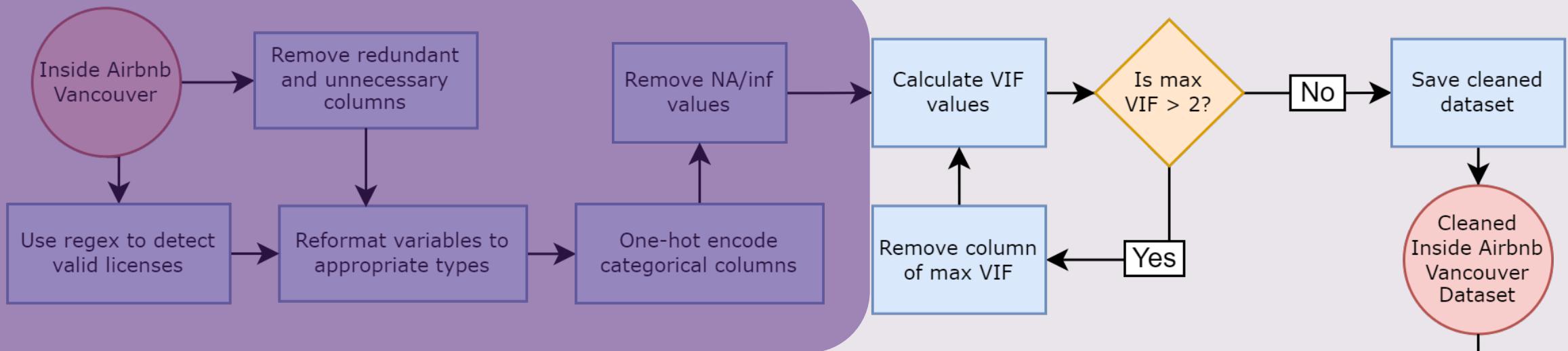
'Review\_Scores': ★4.87/5.00

**Ordinal Data**

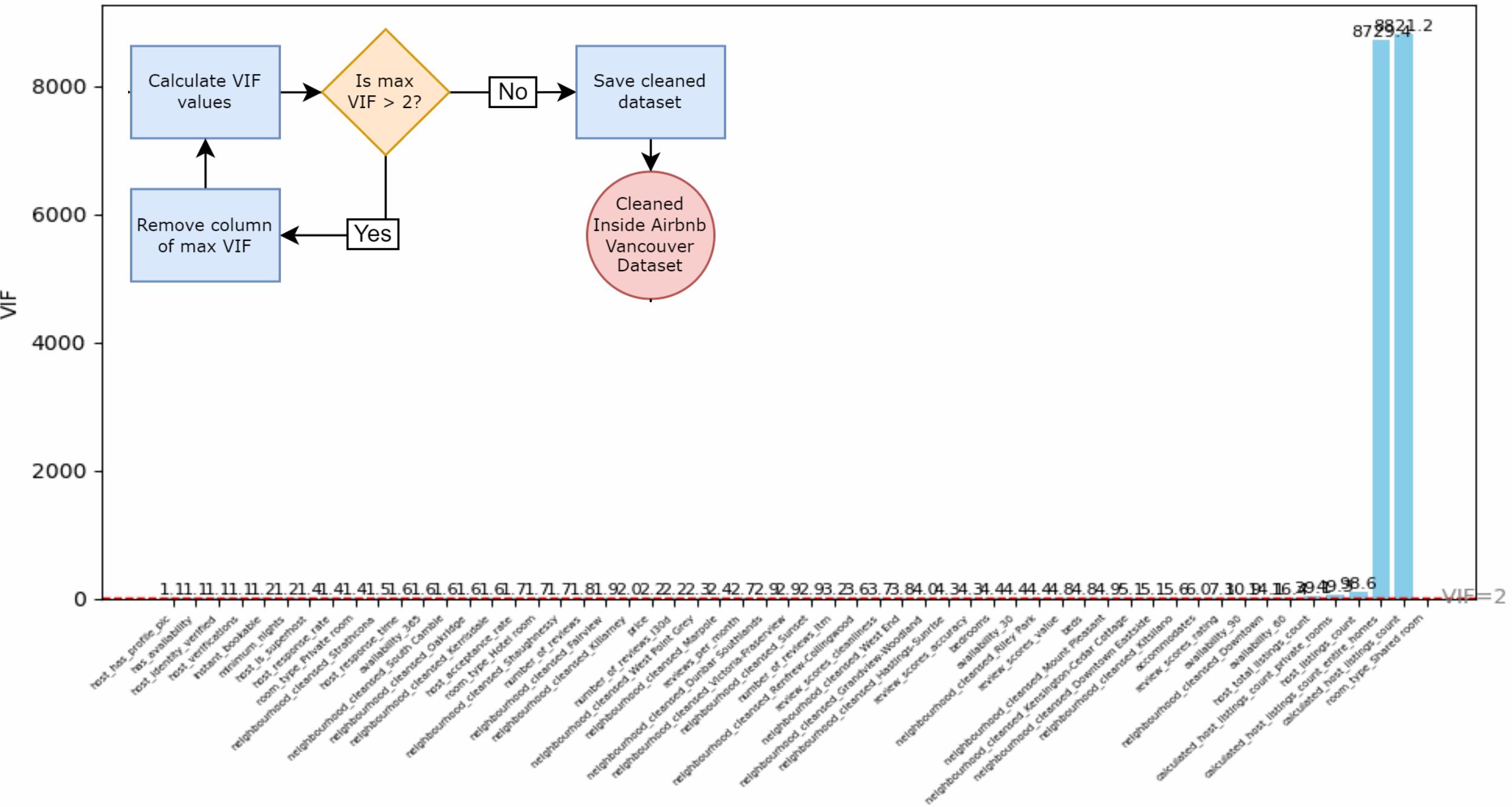
**Categorical Data**

**'Room\_Type'**: <One-Hot coding>

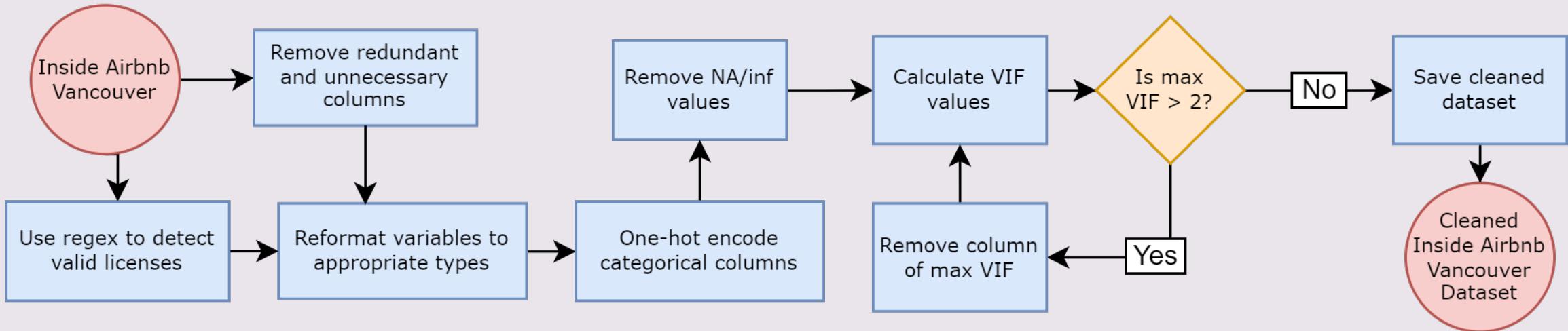
## Preprocessing



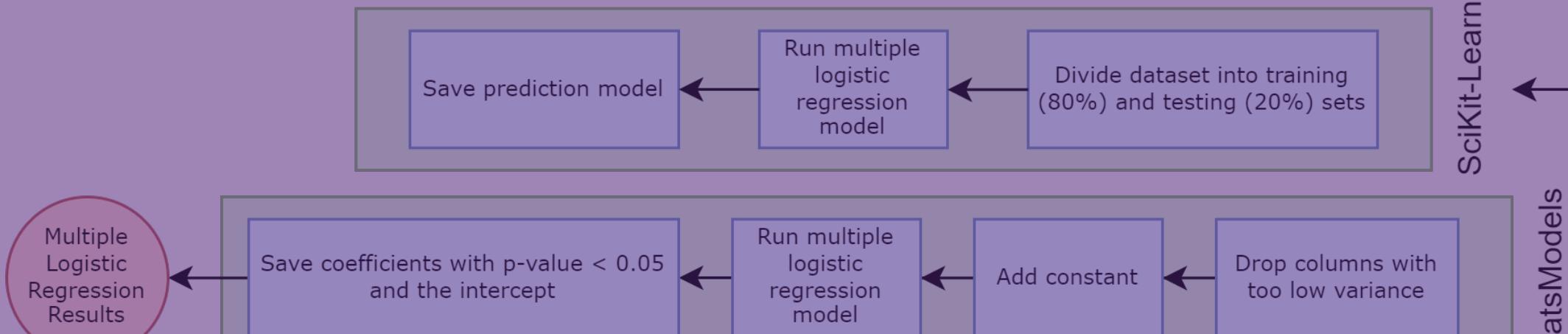
## VIF process



## Preprocessing

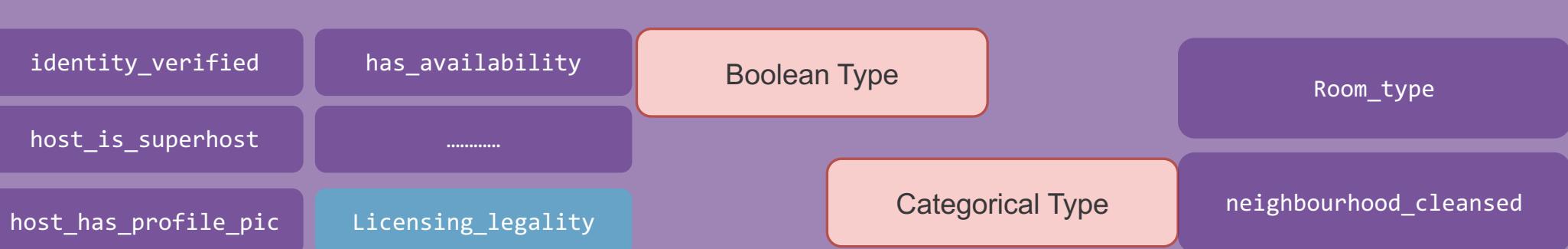
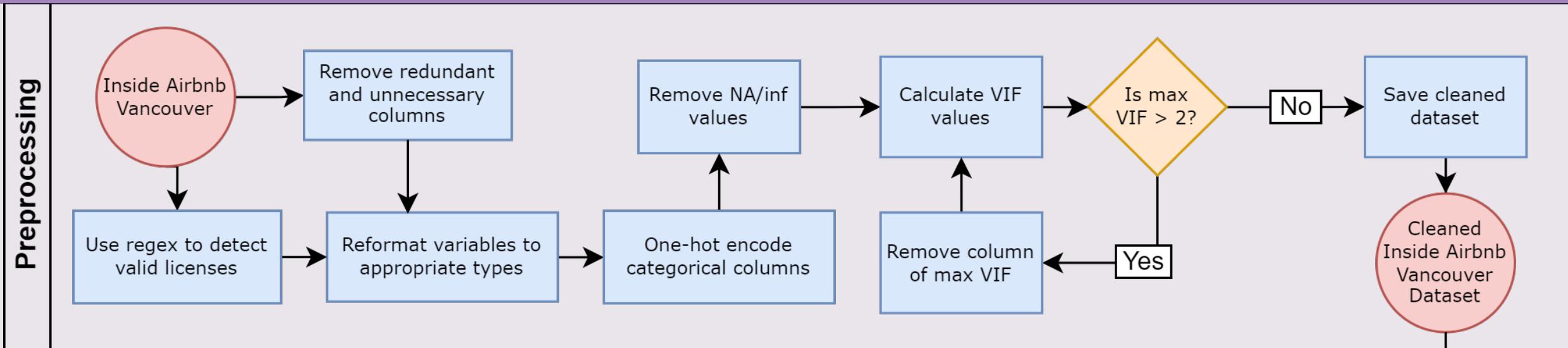
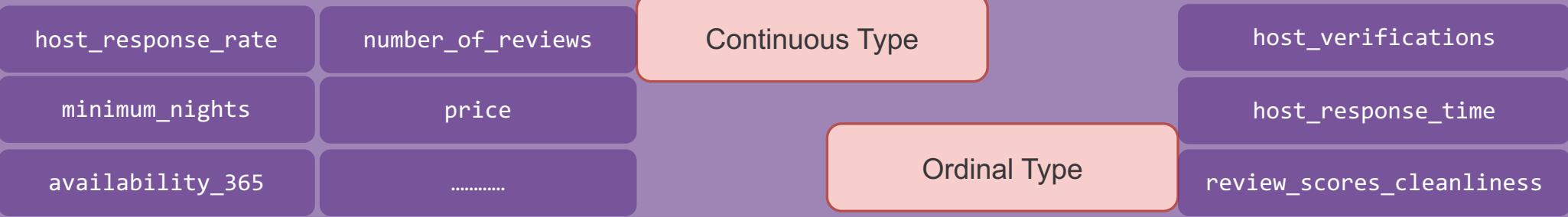


## Modeling



SciKit-Learn

StatsModels



host\_response\_rate

number\_of\_reviews

Continuous Type

host\_verifications

minimum\_nights

price

host\_response\_time

availability\_365

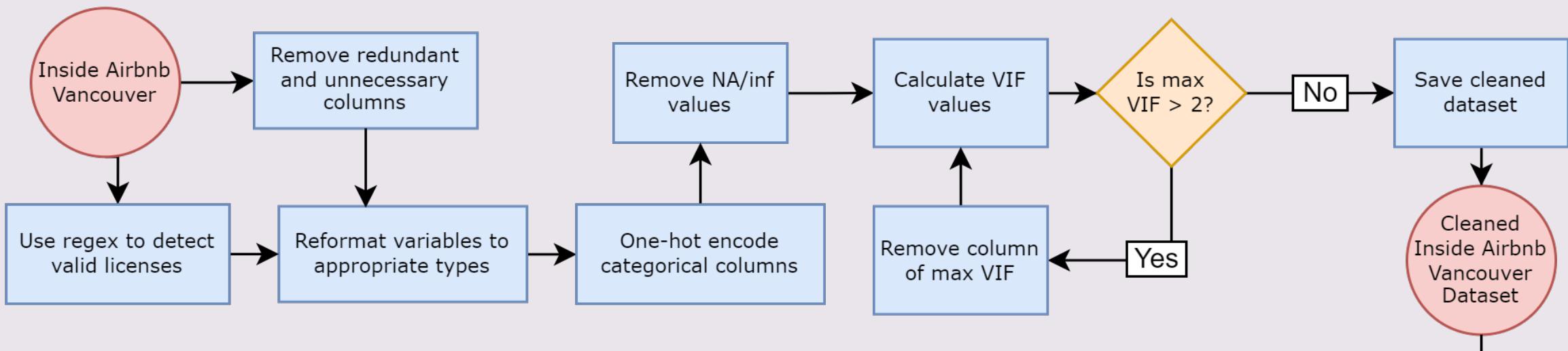
.....

Ordinal Type

review\_scores\_cleanliness

## Normalisation

### Preprocessing



identity\_verified

has\_availability

Boolean Type

Room\_type

host\_is\_superhost

.....

host\_has\_profile\_pic

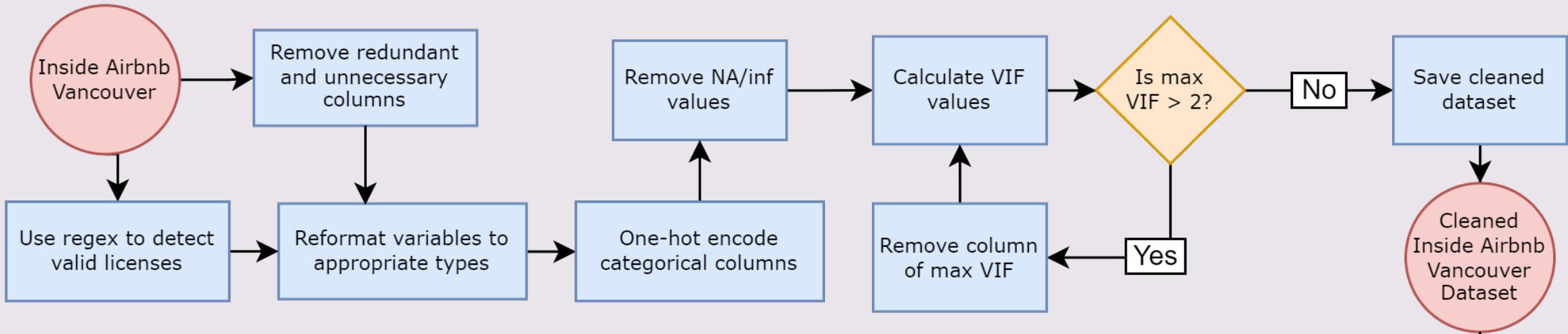
Licensing\_legality

Categorical Type

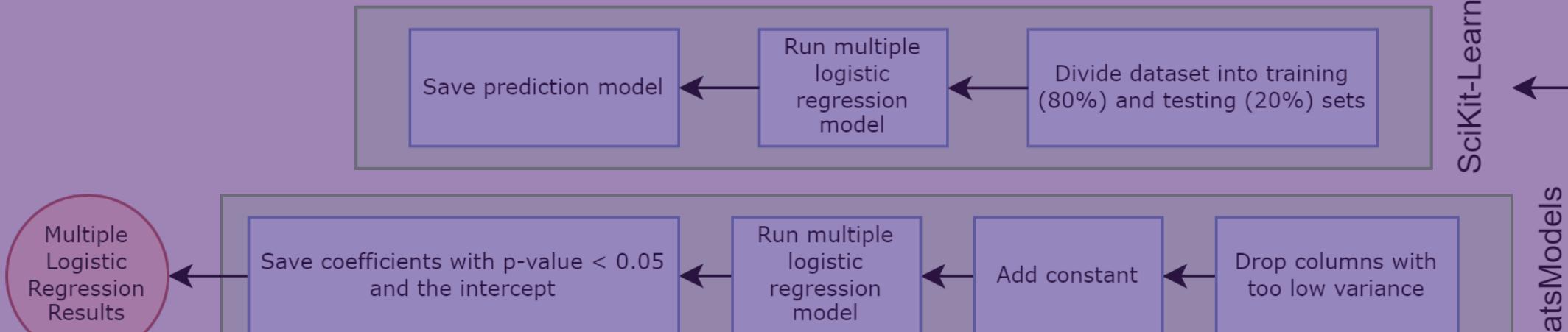
neighbourhood\_cleansed

## Low Variance

## Preprocessing



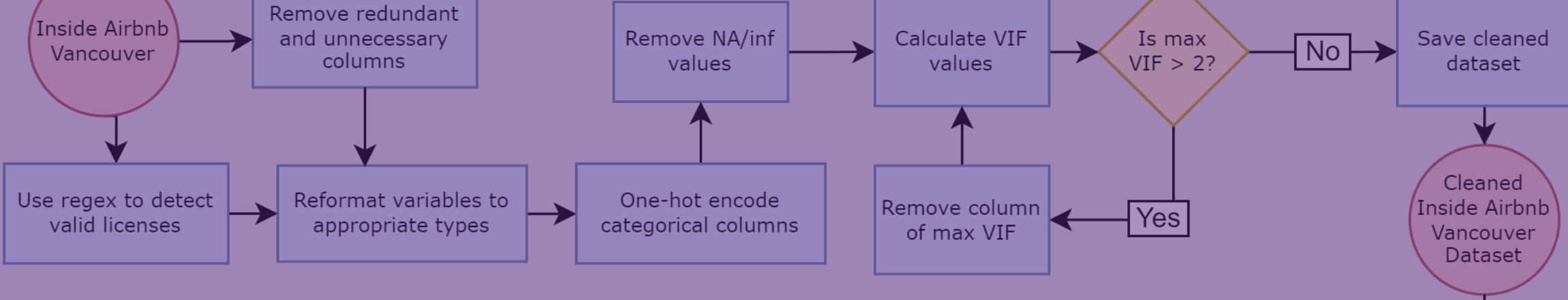
## Modeling



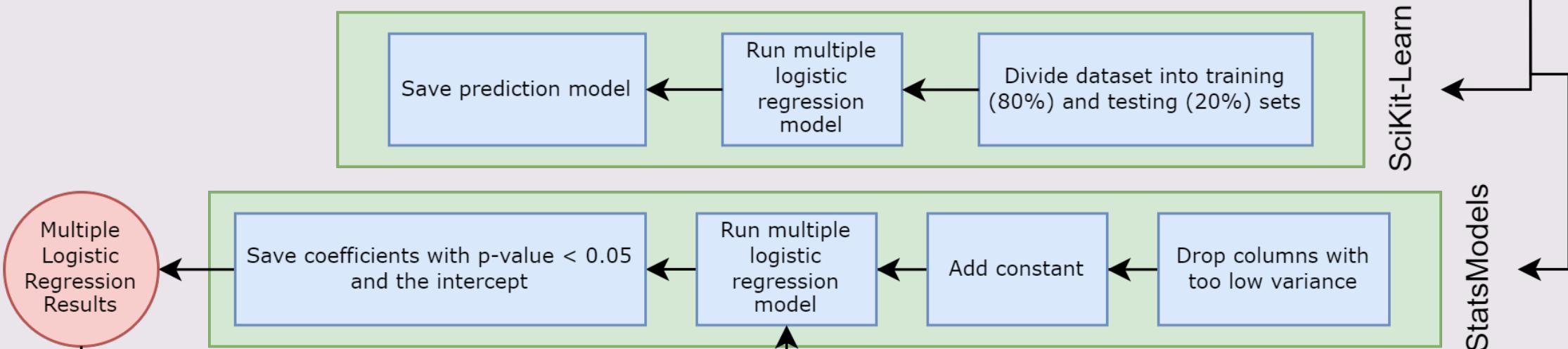
SciKit-Learn

StatsModels

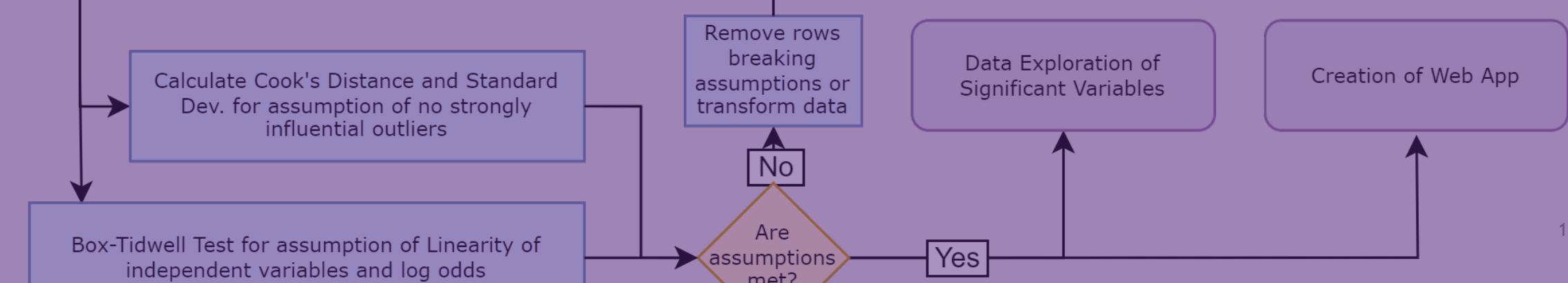
## Preprocessing



## Modeling

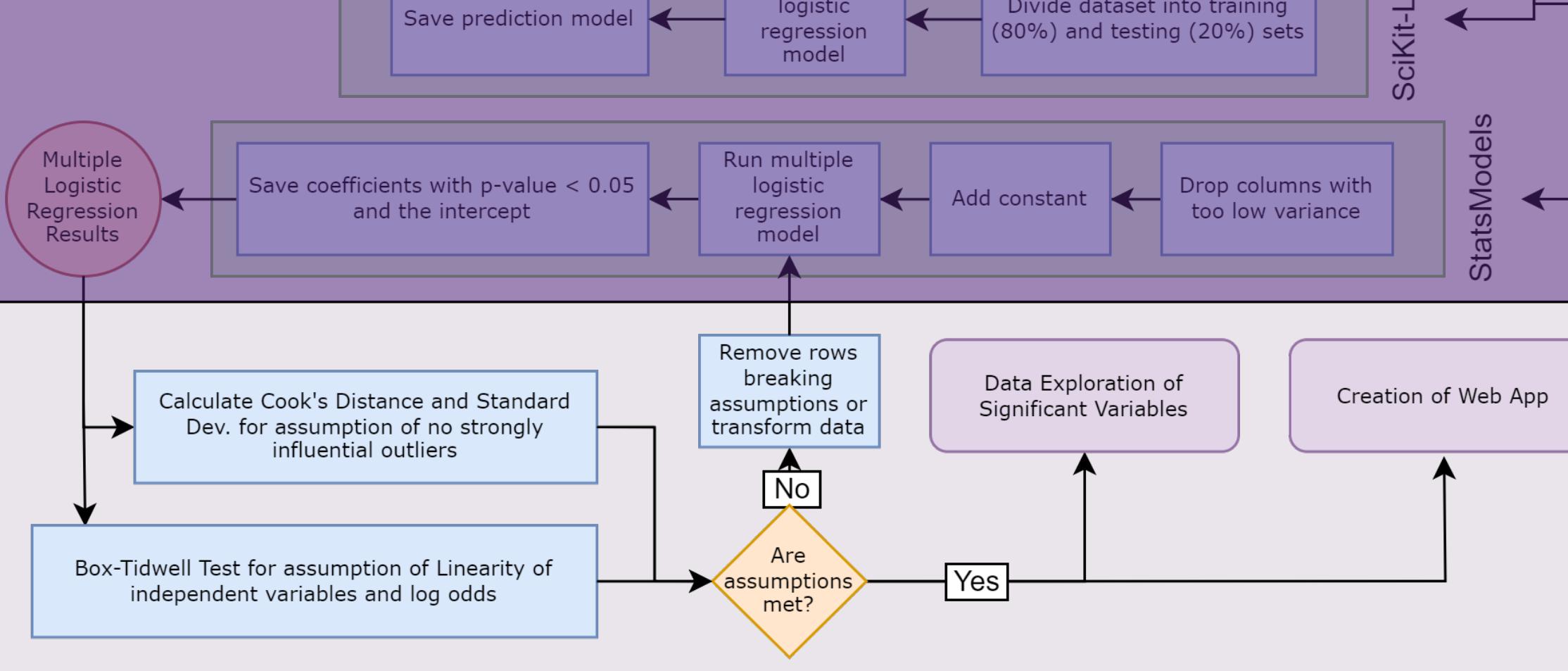


## Validation & Analysis



## Verification & Analysis

## Modeling



# Web App

Full Model:  
90.7% Accuracy

Web App Model:  
85.4% Accuracy



## Short-Term Rental License Probability Calculator: Vancouver

This app uses a multiple logistic regression model to calculate the probability that a short-term rental in Vancouver is licensed or not. The model was trained on data from InsideAirbnb.

**Input the following information about the rental:**

Number of Reviews

17

- +

Price (in CAD)

100.00

- +

Is the rental a private room?

Yes

No

Neighbourhood

Downtown Eastside

**The probability that this rental is licensed is:**

**30.73%**

# Multiple Logistic Regression Model Results

Variable	Coefficient	P-Value
Price	0.0007	0.165
# of Reviews	0.0185	< 0.000
# of Beds	0.2407	0.001
NBHD: Killarney	0.3501	0.538
...	...	...
NBHD: West End	-0.4245	0.059

Variable	Coefficient	P-Value
Price	0.0007	0.165
# of Reviews	0.0185	< 0.000
# of Beds	0.2407	0.001
NBHD: Killarney	0.3501	0.538
...	...	...
NBHD: West End	-0.4245	0.059

Variable	Coefficient	P-Value
# of Reviews	0.0185	< 0.000
# of Beds	0.2407	0.001
...	...	< 0.05

# Assumption 1: No Strongly Influential Outliers

## Influence

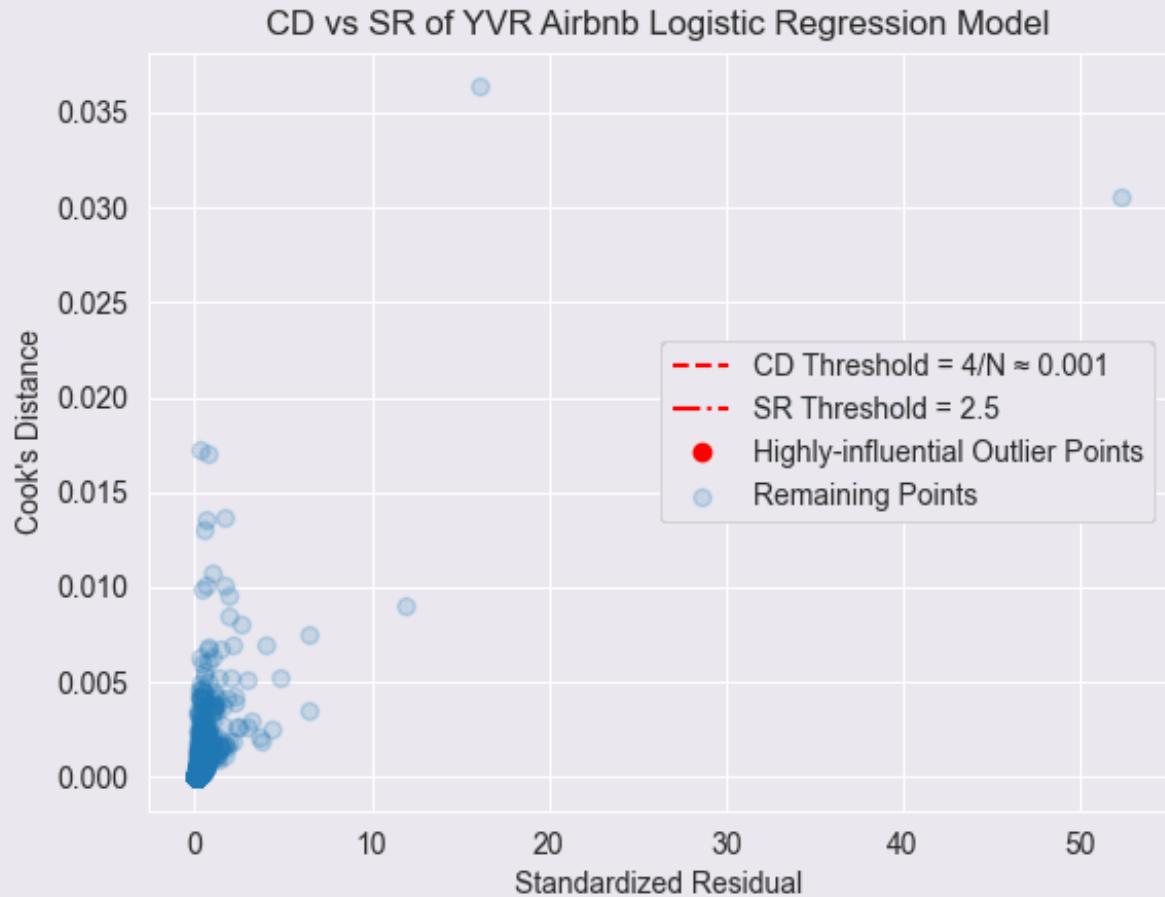
- Using Cook's Distance to determine the influence of a data point, and it is calculated based on its residual and leverage.
- Observations with Cook's Distance  $> 4/N$  are deemed as influential.

## Outliers

- Data points with absolute standardized residual values greater than 2.5 represent possible extreme outliers.

## Putting Both Together

- We can identify the strongly influential outlier data points by finding the top observations based on thresholds defined earlier for Cook's Distance and standardised residuals.
- Then we delete those influential outliers.



# Assumption 2: Linearity of Independent Variables and Log-odds

## Find non-linearity relation

- Assess the scatter plot, There seems no linear relation between price and log-odds.
- For Box-Tidwell test, we expect the p-value of interaction term shall be above 0.05, which indicates the variable holds assumption.
- Applying Box-Tidwell test, using interaction term ( $X \cdot \log X$ ) to re-fit the model. The outcome shows the p-value of interaction term is less than 0.05, so there should be a non-linearity relation.

Box-Tidwell test, Re-fit logit model with interaction term

	coef	std err	z	P> z	[0.025	0.975]
price	0.0221	0.002	13.234	0.000	0.019	0.025
host_response_rate	-0.3835	0.229	-1.678	0.093	-0.831	0.064
review_scores_checkin	5.6933	1.751	3.252	0.001	2.261	9.125
review_scores_cleanliness	1.0826	1.828	0.592	0.554	-2.501	4.666
price:Log_price	-0.0024	0.000	-12.542	0.000	-0.003	-0.002

Scatter plot of price and log-odds



# Assumption 2: Linearity of Independent Variables and Log-odds

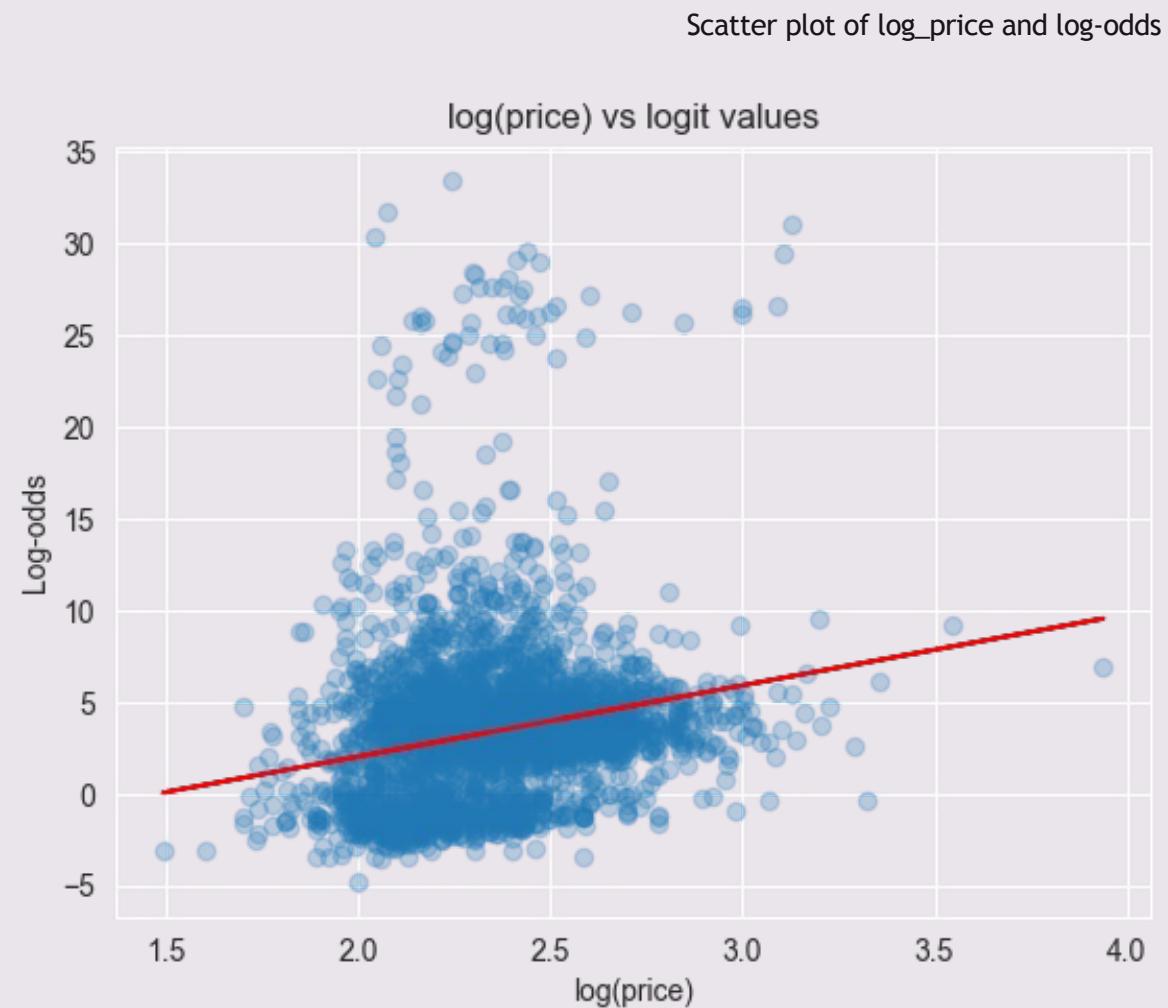
## Transforming “price” variable

- We transform the “price” to "log price", Plotting the scatter between log(price) and log-odds, visually check they have a more reliable linear relation now.

## Re-do the Box-Tidwell Test

- Applying Box-Tidwell test to the transforming variable log\_price, The p-value of interaction term is 0.069, more than 0.05, indicating that the transforming variable holds assumption.

The second Box-Tidwell test						
	coef	std err	z	P> z	[0.025	0.975]
const	-20.8641	7.543	-2.766	0.006	-35.647	-6.081
log_price	8.5375	3.815	2.238	0.025	1.060	16.015
log_price:Log_log_price	-2.6027	1.432	-1.818	0.069	-5.409	0.204



# Data Exploration

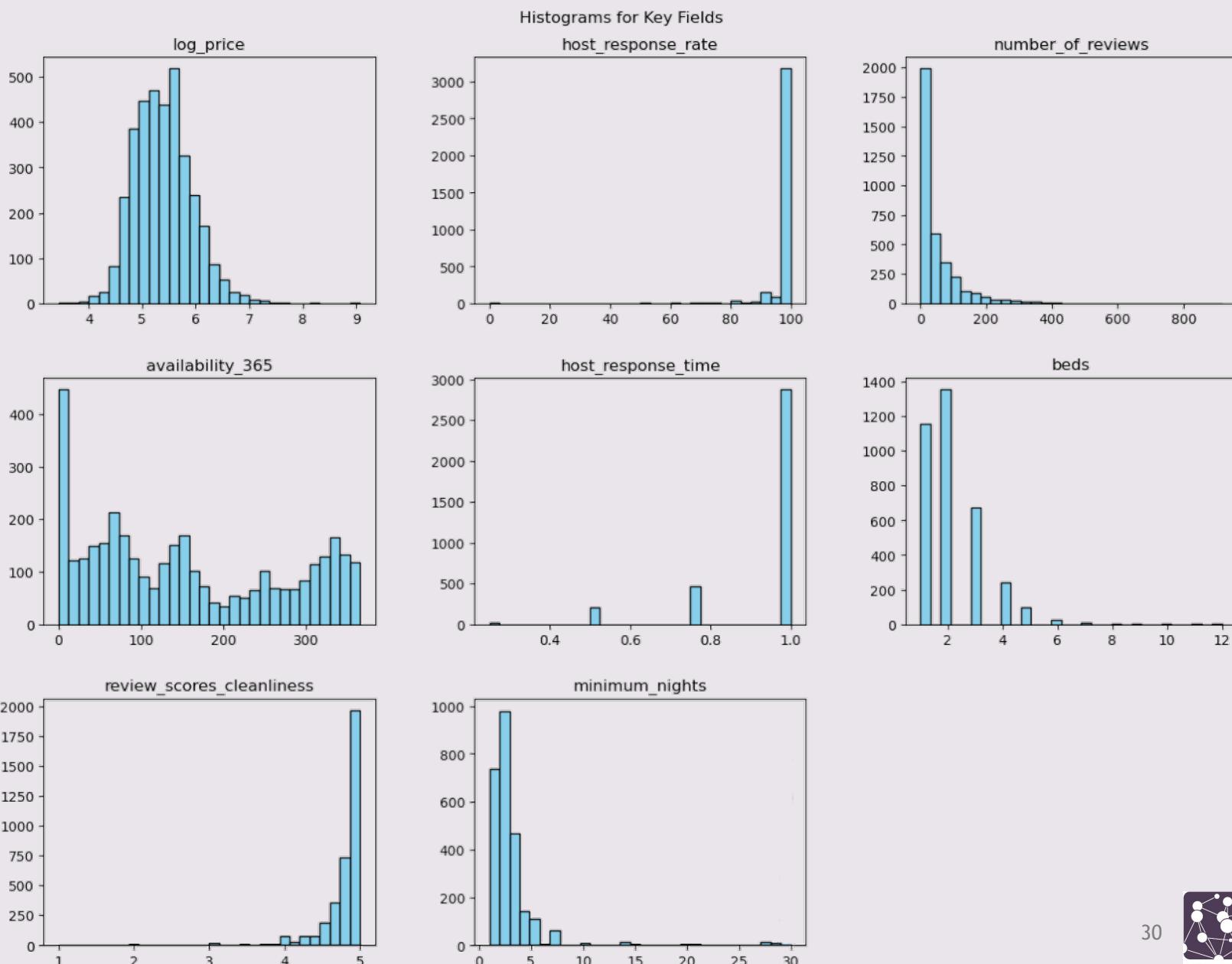
# The List of Key Variables

Columns	Data Type	Columns	Data Type
Log_price	float64	room_type_Private room	binary
host_response_rate	float64	neighbourhood_cleansed_Grandview-Woodland	binary
number_of_reviews	float64	neighbourhood_cleansed_Hastings-Sunrise	binary
availability_365	float64	neighbourhood_cleansed_Kensington-Cedar Cottage	binary
host_response_time	float64	neighbourhood_cleansed_Mount Pleasant	binary
review_scores_cleanliness	float64	neighbourhood_cleansed_Riley Park	binary
minimum_nights	float64	neighbourhood_cleansed_Strathcona	binary
beds	categorical	neighbourhood_cleansed_Victoria-Fraserview	binary
		neighbourhood_cleansed_West Point Grey	binary
		legal_listing	binary

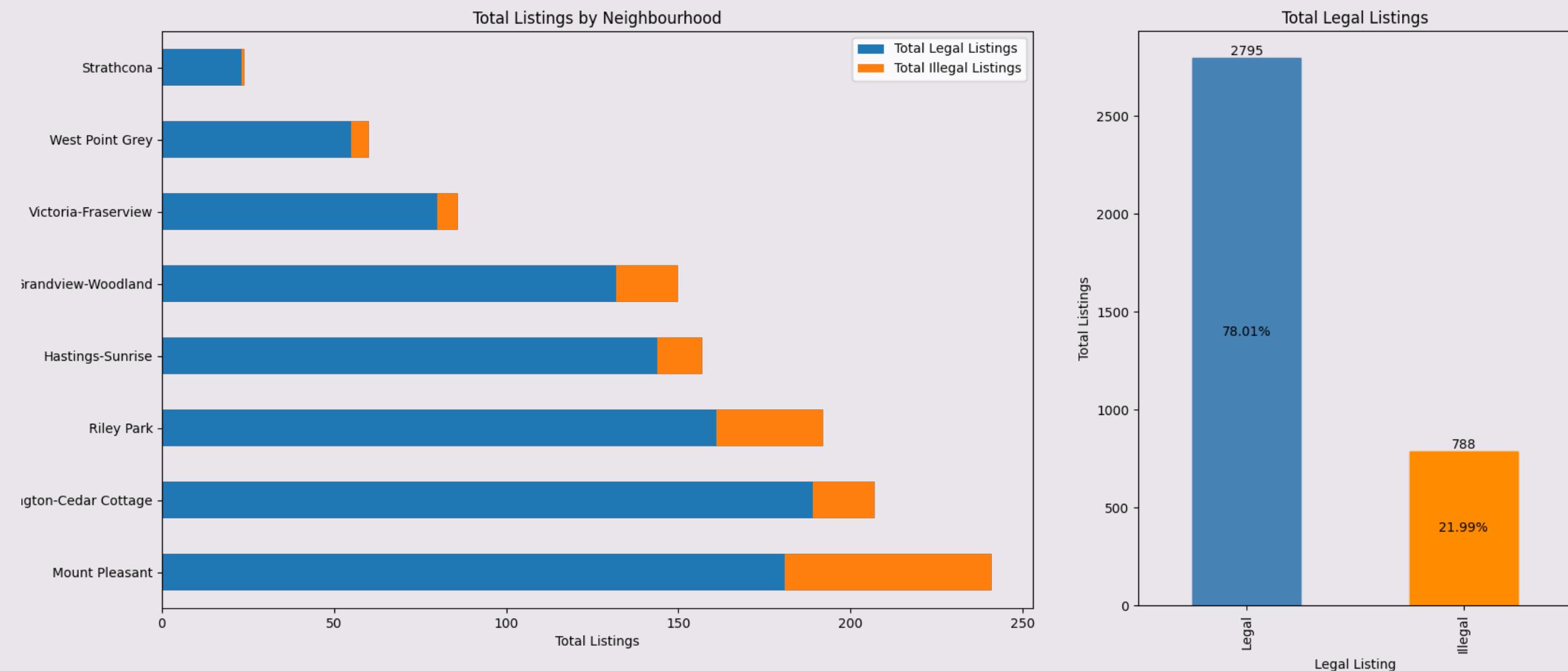
# Histograms for Key Variables

## Float and categorical variables

- The transforming variable 'log\_price' has a "normal distribution".
- The availability\_365 seems to have a "uniform distribution". Others have skewness.
- Only minimum\_nights Less than 30 which will be required to have a license in Vancouver.
- So, our dataset only include data below 30.

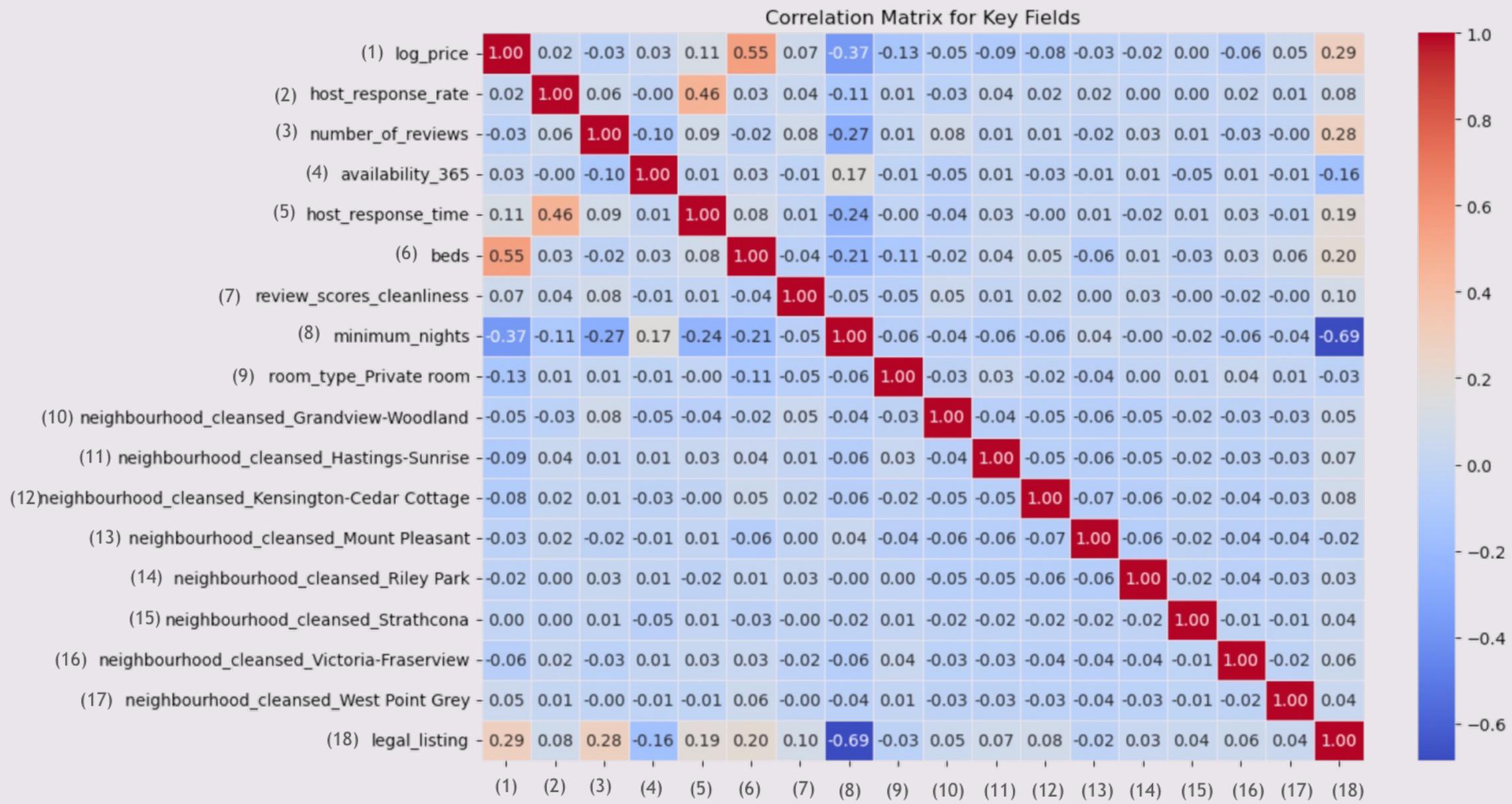


# Stacked Bar Plot for “neighbourhood” Variables



Shows the proportion of legality of listings in different neighbourhoods

# Correlation Matrix for Key Variables

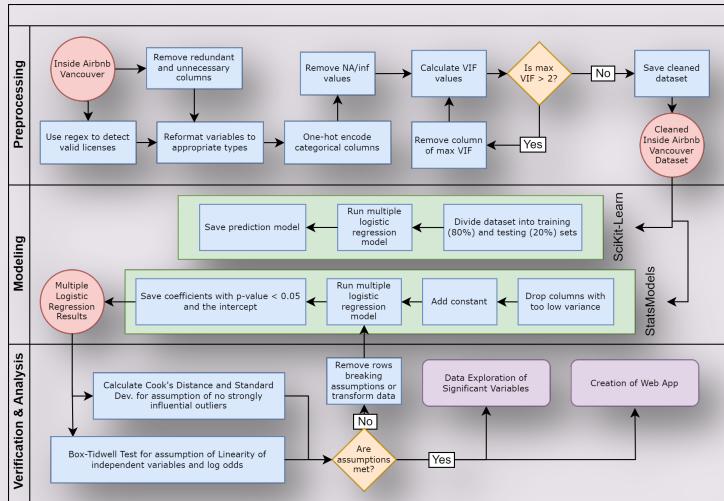


A B C

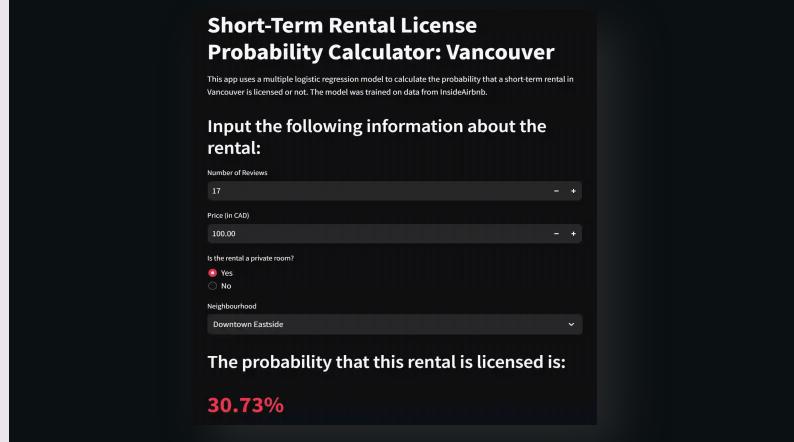
# Conclusion

# In Summary...

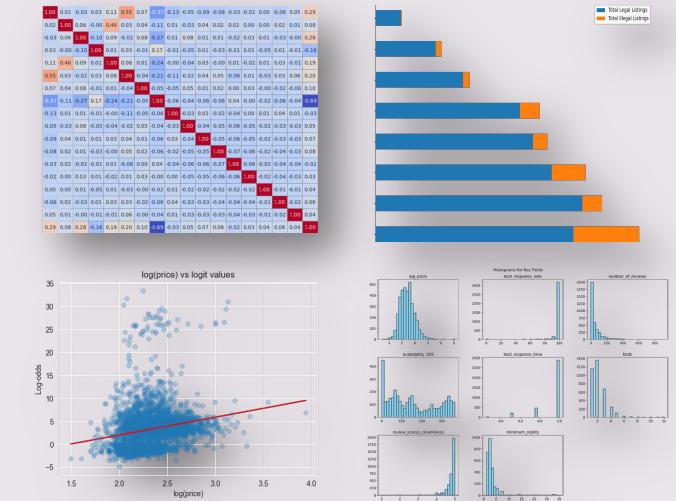
## A Data Processing Cleaning

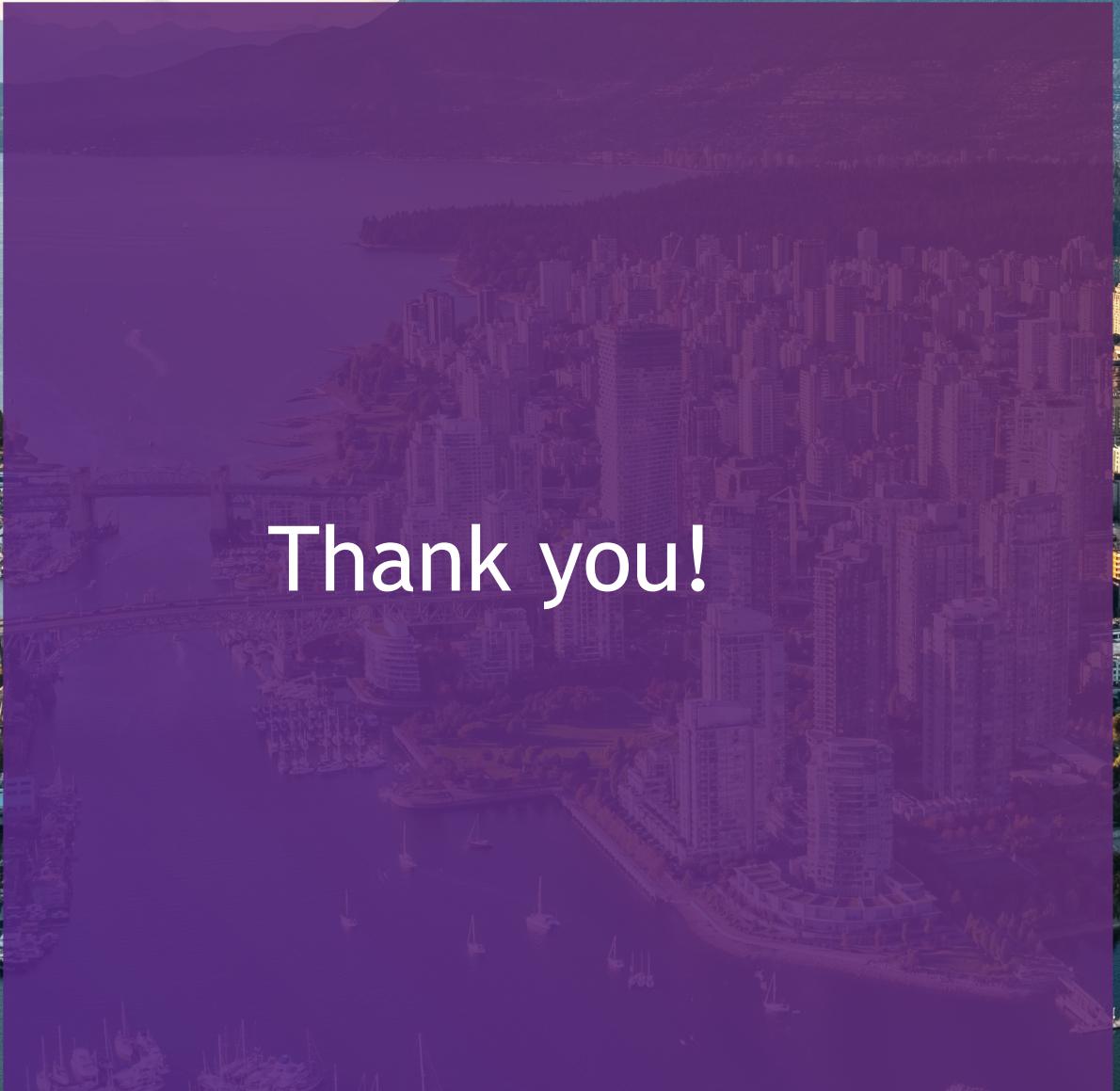


## B Modelling to Identify Key Fields



## C Data Exploration





Thank you!





Github Repository



Web App