

Understanding the Impact of Airbnb on Gentrification & Homelessness

In the City of Toronto

Angela Huang

April 2022

Business Case

Background

- Homelessness is an urgent and growing issue in Toronto.
- 94 per cent of those experiencing homelessness in Toronto want permanent housing, but face barriers in securing it.
- In the past 10 years, average market rent for a one-bedroom has increased by 33 per cent.
- There are thousands of homes that could be on the long-term rental market but aren't.
- Many of these are owned by absentee investors who are leaving them vacant for their eventual use, but many more are owned by Airbnb hosts who can make more on Short-Term rentals than they could renting long-term to a family.

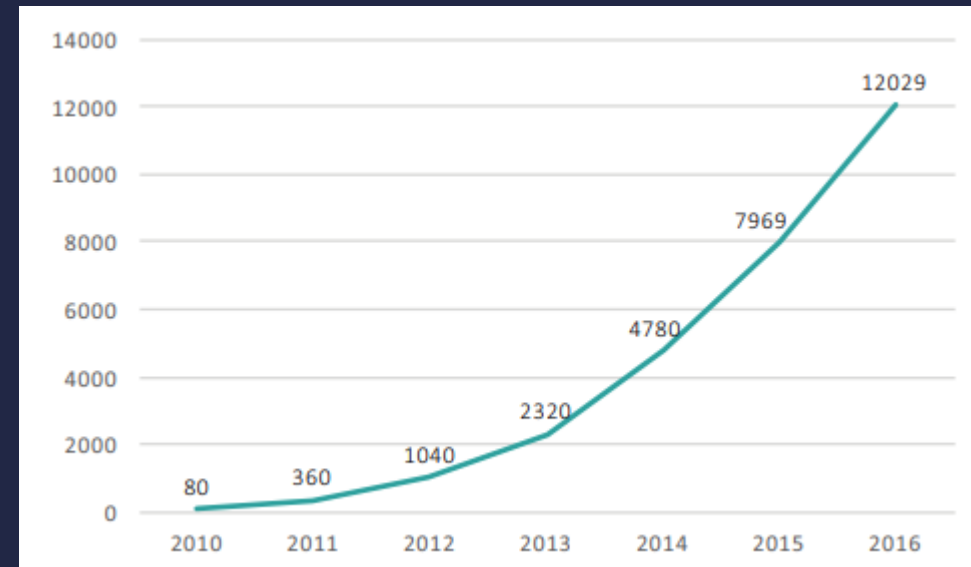
Toronto Rental Market

Average Rental Vacancy Rates in Toronto, 2006 to 2018



Source: Canada Mortgage and Housing Corporation, Rental Market Survey

Total Number of Airbnb Listings in Toronto, 2010 to 2016



Source: Slee and Custom Tabulations

Objective



Understanding the effect of Short-Term rentals on Homelessness



Create a machine learning model to make these predictions more data-driven and accurate vs traditional human data modeling



The model would improve regulatory actions and shed light on the effectiveness of such regulations by policy makers

Allows policy makers to prioritize which areas to investigate for potential non-compliant Short-Term rental operators
Identifies which regulatory decisions may have the most beneficial impact

Data

The City of Toronto collects data on:

- Daily Shelter & Overnight Service Occupancy & Capacity
 - SERVICE_USER_COUNT: Count of the number of service users staying in an overnight program as of the occupancy time and date.
- Daily Registered Short-Term Rentals Operators

Inside Airbnb, a mission driven activist project, collects data on:

- Quarterly scrape of publicly available listing information from Airbnb's site
 - Daily Price
 - Minimum number of nights
 - Availability (either booked or blocked)

Requirements & Benefits of Using Cloud

- **Scalable** - Support for increased data sources, or increased compute power on demand
- **Cost effective** - Less upfront cost (Low CAPEX)
- **Secure** - Can set access permissions and support encryption
- **Low Maintenance** - Reliable - Have failover and additional servers
- **Setup quickly (initial POC)** - Without requiring too much time to configure and acquire equipment

Solution Architecture

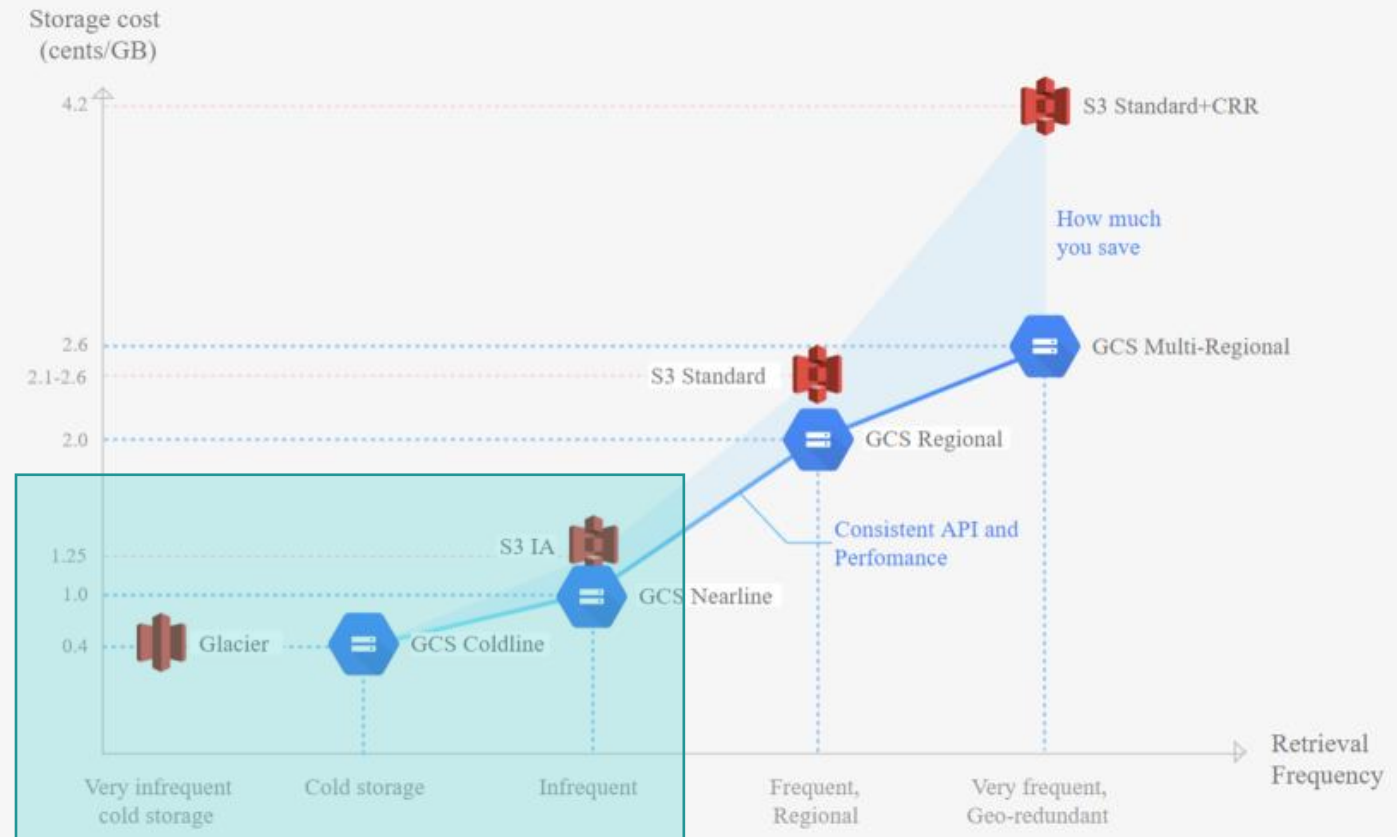
Functionality & Assumptions

| Functionality | Assumptions |
|--|---|
| Store: Raw data | <ul style="list-style-type: none">• Annual additional storage of 1GB – at most, data is accessed once a year |
| Compute: Upload data onto database/data warehouse | <ul style="list-style-type: none">• Automatically load data into database once file is dropped into storage (runs daily or quarterly) |
| Store & Transform/Process: Data clean up, aggregation, feature engineering | <ul style="list-style-type: none">• Automatically process (updates daily or quarterly)• Relational Database |
| Create: Machine Learning model | <ul style="list-style-type: none">• Limited staff, ideally an explainable model with very little knowledge in ML |

Storage: AWS vs. GCP

AWS S3 Bucket is more desirable option

Infrequent Retrieval since it will be stored in a database/ data warehouse



Compute: AWS vs. GCP

| | AWS Lambda | GCP Cloud Functions |
|--|---|--|
| Pricing: <ul style="list-style-type: none">Free Monthly Duration (GB-seconds)Free Monthly RequestsCost of Each Additional 1 Million RequestsCost of Each Additional 1 GB-second | <ul style="list-style-type: none">400,0001 Million\$0.20\$0.000016 | <ul style="list-style-type: none">400,0002 Million\$0.40\$0.0000125 |
| Execution Time - Maximum Timeout | 15 minutes | 9 minutes |
| Memory | 128 MB – 10240 MB | 128 MB – 4096 MB (in multiples of 128 MB) |

AWS Lambda is more desirable option - More execution time and memory

Database: AWS vs. GCP

| | AWS Redshift | GCP BigQuery |
|---------|---|---|
| Pricing | <ul style="list-style-type: none">• Storage: \$306 per TB per month• Queries: unlimited processing• Pricing by on-demand/by-the-hour nature | <ul style="list-style-type: none">• Storage: \$20 per TB per month• Queries: \$5/TB• Pricing by query |

3 Key Differences of RedShift vs. BigQuery

1. Amazon RedShift is provisioned on clusters and nodes. Google BigQuery is serverless.
2. RedShift supports 1,600 columns in a single table, BigQuery supports 10,000 columns.
3. RedShift requires periodic management tasks like vacuuming tables, BigQuery has automatic management.

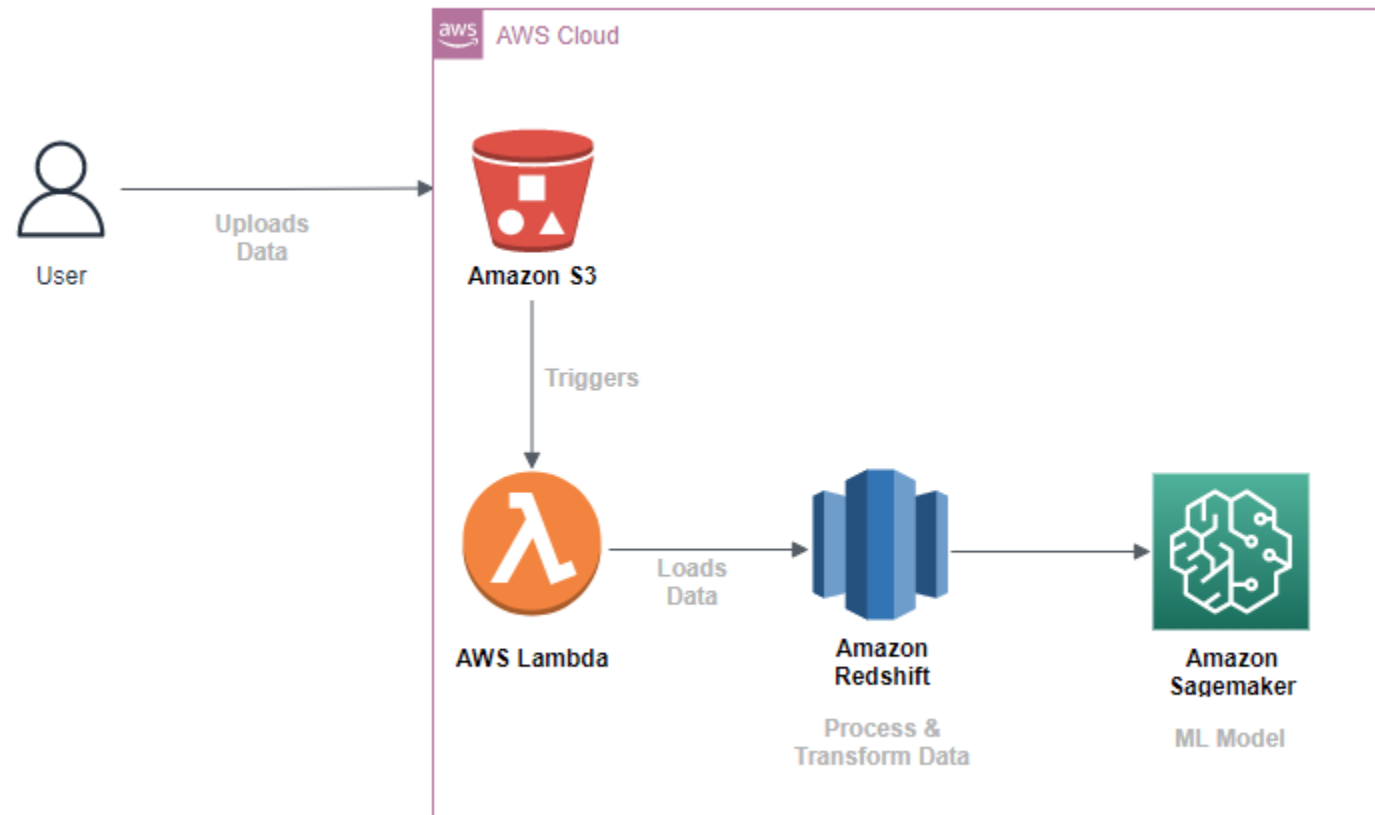
GCP BigQuery is more desirable option –
Variant workload

Machine Learning: AWS vs. GCP

- AWS: Amazon Sagemaker Autopilot
- GCP: Google Cloud AutoML

Both offers very similar capabilities

AWS Solution Architecture



Implementation of Architecture

Amazon S3

Amazon S3 > Buckets > 3760-term-project

3760-term-project [Info](#)

[Objects](#) | [Properties](#) | [Permissions](#) | [Metrics](#) | [Management](#) | [Access Points](#)

Objects (5)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

[Refresh](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#)


[Actions](#) [Create folder](#) [Upload](#)


< 1 > [Settings](#)


| <input type="checkbox"/> | Name | Type | Last modified | Size | Storage class |
|--------------------------|---|--------|---------------|------|---------------|
| <input type="checkbox"/> | final dataframe/ | Folder | - | - | - |
| <input type="checkbox"/> | Inside Airbnb Listings/ | Folder | - | - | - |
| <input type="checkbox"/> | ML Results/ | Folder | - | - | - |
| <input type="checkbox"/> | Shelter & Overnight Capacity/ | Folder | - | - | - |
| <input type="checkbox"/> | Short Term Rental Registration/ | Folder | - | - | - |

AWS Lambda

Function overview [Info](#)

 data-ingestion-3-7

 Layers (2)

 S3

[+ Add trigger](#)

[+ Add destination](#)

Description

-

Last modified

1 day ago

Function ARN

[arn:aws:lambda:us-east-1:619780975338:function:data-ingestion-3-7](#)

Function URL [Info](#)

-

[Code](#) | [Test](#) | [Monitor](#) | [Configuration](#) | [Aliases](#) | [Versions](#)

Code source [Info](#)

[Upload from](#)

File Edit Find View Go Tools Window [Test](#) [Deploy](#)

Go to Anything (Ctrl-P)

Environment

data-ingestion-3-7

lambda_function.py

```
1 import boto3
2 import json
3 import psycpg2
4
5 s3_client = boto3.client('s3')
6 redshiftdb_client = boto3.client('redshift')
7
8 def lambda_handler(event, context):
9     # TODO implement
10
11     # get bucket name
12     bucket = event['Records'][0]['s3']['bucket']['name']
13     # get file name
14     file_name = event['Records'][0]['s3']['object']['key']
15
16     s3_object = s3_client.object(Bucket=bucket, Key=file_name)
17
18     s3FileReader = s3_object['Body'].read()
19     s3Dict = json.loads(s3FileReader)
20
```


Amazon Redshift

Amazon SageMaker Autopilot

File Edit View Run Kernel Git Tabs Settings Help

Launcher term-project-3760 Linear Regression

AUTOPILOT JOB 33 minutes ago Deploy model

term-project-3760

Problem type: Regression Open candidate generation notebook Open data exploration notebook

Best model endpoint: Sagemaker-endpoint

Trials Job profile

0 rows selected

| Best Model | Model name | Status | Start time | Objective: Mse |
|------------|---------------------------------|-----------|-------------|----------------|
| Best Model | term-project-3760libZHO67oBs... | Completed | 1 hour ago | 0.428 |
| | term-project-3760libZHO67oBs... | Completed | 2 hours ago | 57622.867 |
| | term-project-3760libZHO67oBs... | Completed | 2 hours ago | 57244.523 |
| | term-project-3760libZHO67oBs... | Completed | 2 hours ago | 57136.93 |
| | | | | 26272.715 |
| | | | | 20459.502 |
| | | | | 16404.996 |
| | | | | 15369.82 |
| | | | | 11688.222 |

Amazon S3 Buckets 3760-term-project ML Results/ term-project-3760/ Copy S3 URI

term-project-3760/ Objects Properties

Objects (7)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Create folder Upload

Find objects by prefix

| | Name | Type | Last modified | Size | Storage class |
|--------------------------|------------------------------|--------|---------------|------|---------------|
| <input type="checkbox"/> | data-processor-models/ | Folder | - | - | - |
| <input type="checkbox"/> | documentation/ | Folder | - | - | - |
| <input type="checkbox"/> | preprocessed-data/ | Folder | - | - | - |
| <input type="checkbox"/> | sagemaker-automl-candidates/ | Folder | - | - | - |
| <input type="checkbox"/> | transformed-data/ | Folder | - | - | - |
| <input type="checkbox"/> | tuning/ | Folder | - | - | - |
| <input type="checkbox"/> | validations/ | Folder | - | - | - |

Amazon Redshift > Query editor

Editor Query history Saved queries Scheduled queries

Resources Info

Select database Info

To view schemas, select a database.

term_project_db

Select schema Info

To view tables, select a schema.

public

Filter tables

listings

id
name
host_id
host_name
neighbourhood_group
neighbourhood
latitude
longitude
room_type
price
minimum_nights
number_of_reviews

Status Connected database term_project_db user

Change connection

creating s... copy listin...

```
1 CREATE TABLE listings
2 (
3   id
4   name
5   host_id
6   host_name
7   neighbourhood_group
8   neighbourhood
9   latitude
10  longitude
11  room_type
12  price
13  minimum_nights
14  number_of_reviews
15  last_review
```

Run Save Schedule Clear Send feedback

Query results Table details

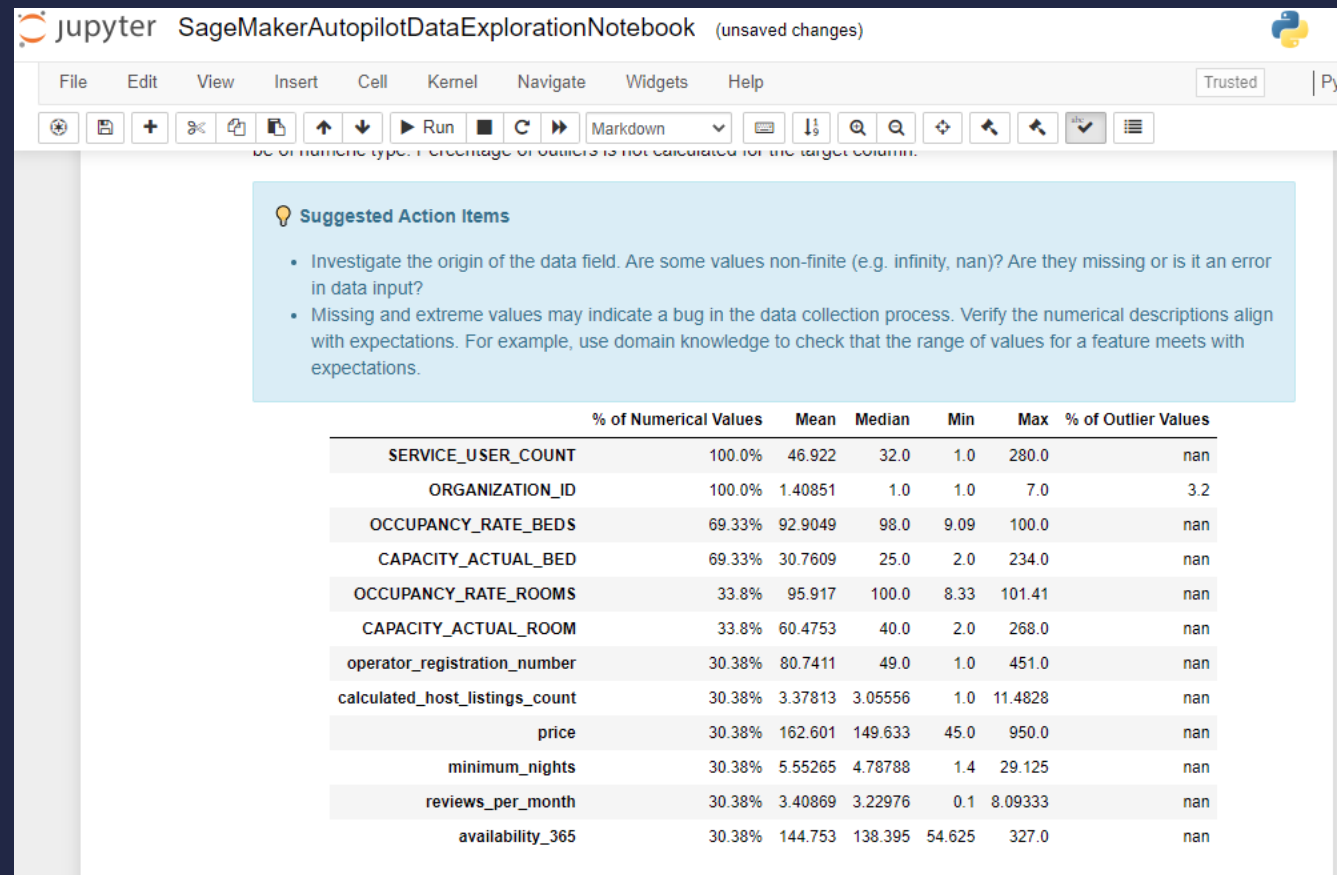
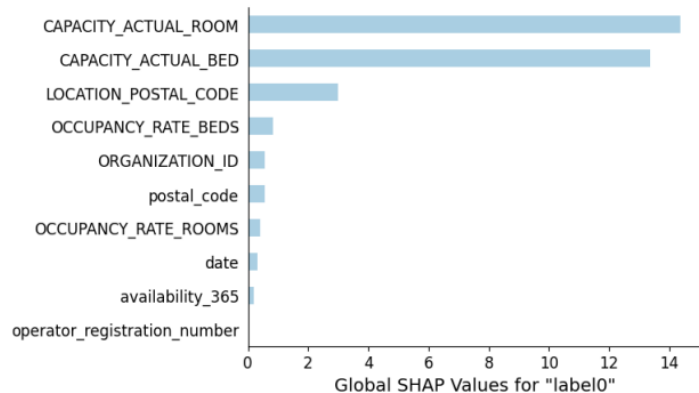
Amazon SageMaker Autopilot - Results

Explainability Report

We report the following SageMaker analysis.

Explanations for "label0"

The Model has 14 input features. We computed KernelShap on the dataset 'dataset' and display the 10 features with the greatest feature attribution.



Challenges

Challenges

- City of Toronto's API only provided first 100 records
- Attaching layer to lambda
- Creating IAM Roles for Redshift