Understanding the Impact of Airbnb on Gentrification & Homelessness

In the City of Toronto

Angela Huang
April 2022

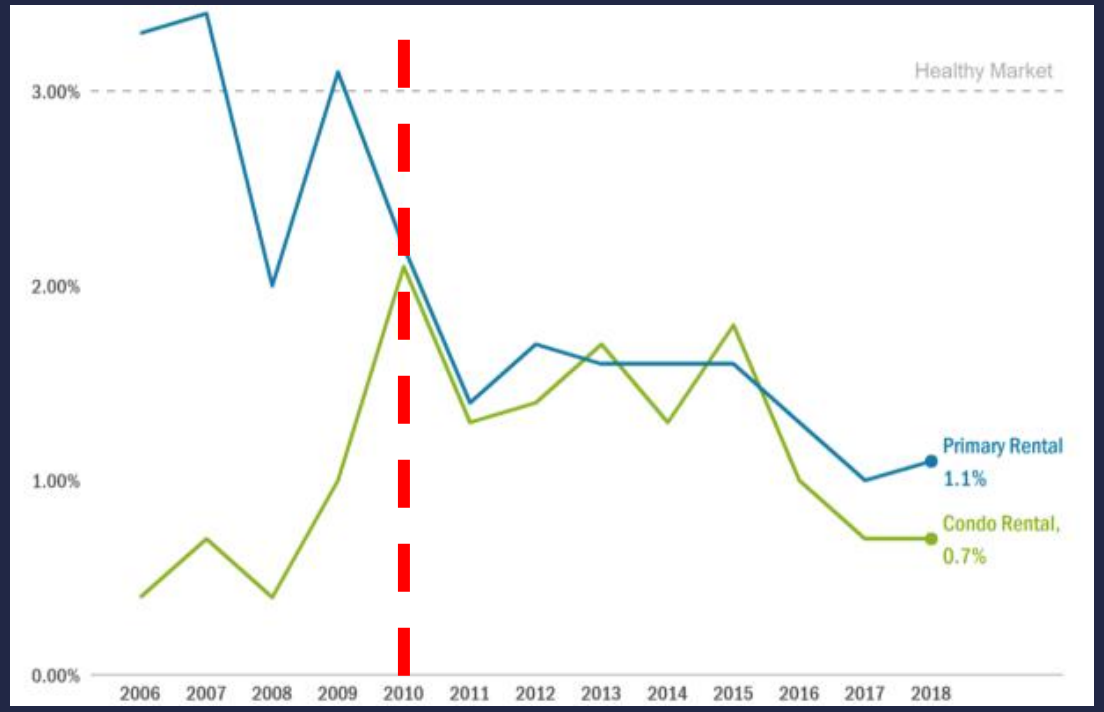
Business Case

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.

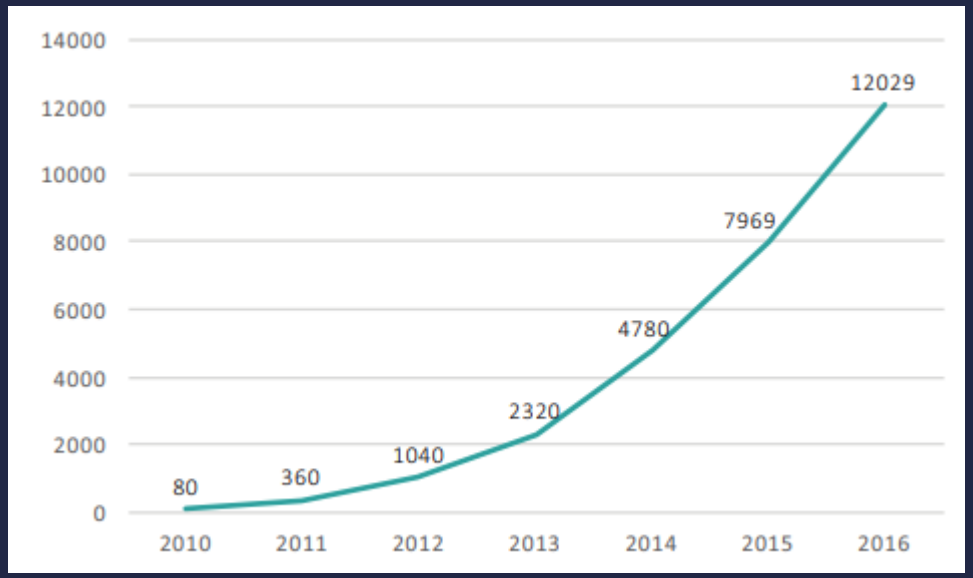
Toronto Rental Market

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

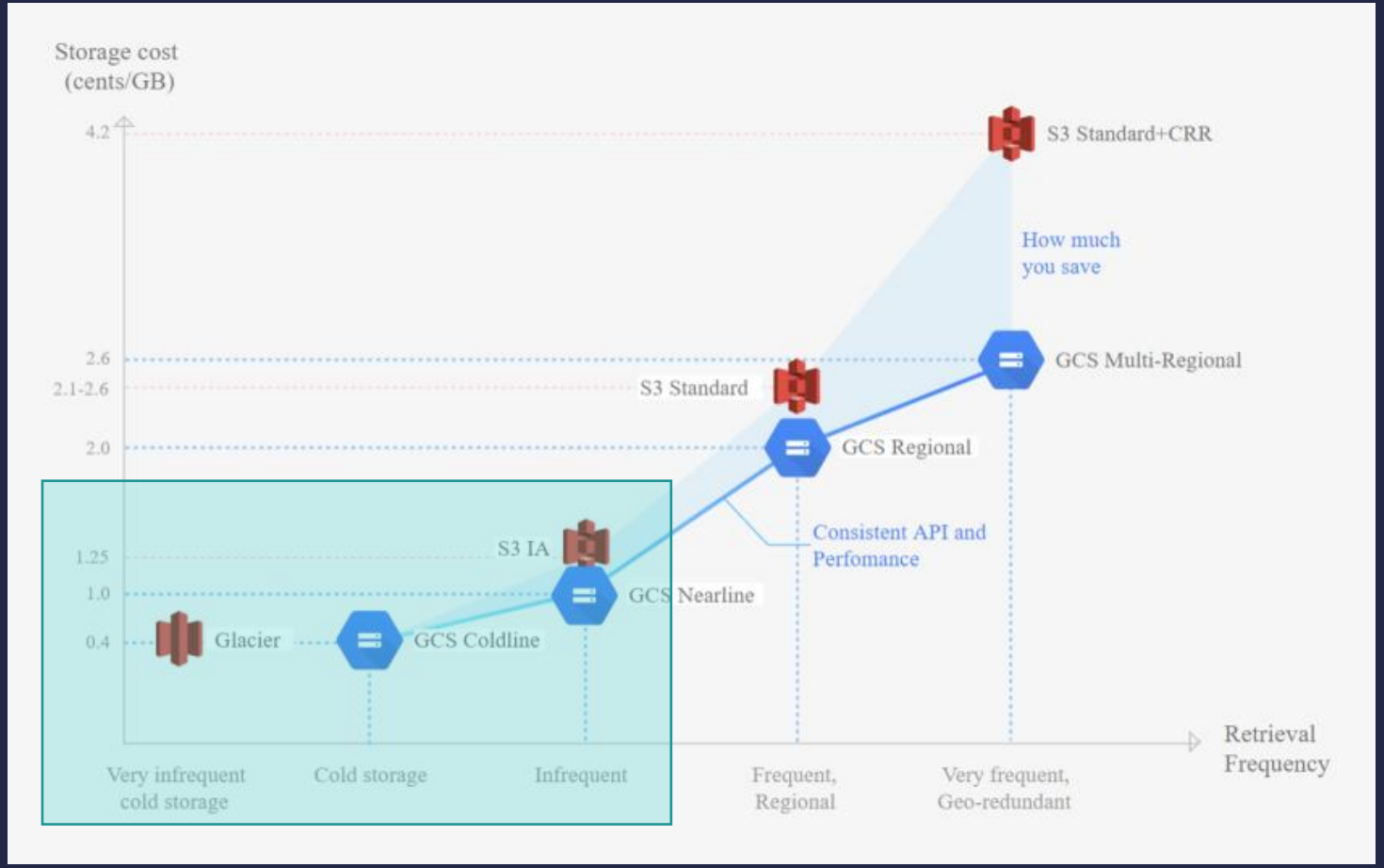
Functionality & Assumptions

Functionality	Assumptions
Store: Raw data	• Annual additional storage of 1GB – at most, data is accessed once a year
Compute: Upload data onto database/data warehouse	• Automatically load data into database once file it dropped into storage (runs daily or quarterly)
Store & Transform/Process: Data clean up, aggregation, feature engineering	• Automatically process (updates daily or quarterly) • Relational Database
Create: Machine Learning model	• 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: • Free Monthly Duration (GB-seconds) • Free Monthly Requests • Cost of Each Additional 1 Million Requests • Cost of Each Additional 1 GB-second	 • 400,000 • 1 Million • $0.20 • $0.000016	 • 400,000 • 2 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	• Storage: $306 per TB per month • Queries: unlimited processing • Pricing by on-demand/by-the-hour nature	• 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.

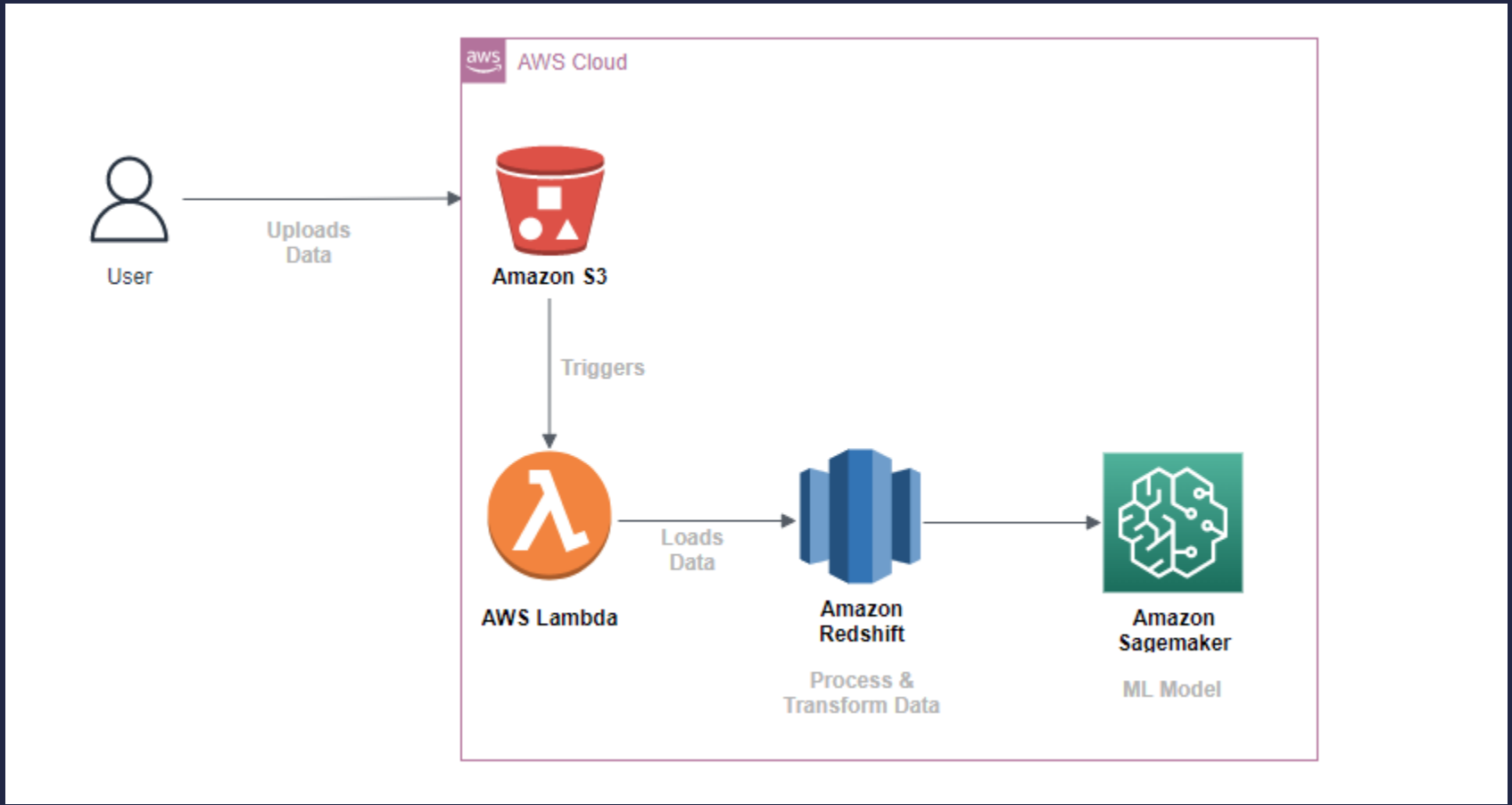
GCP BigQuery is more desirable option – Variant workload

Machine Learning: AWS vs. GCP

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

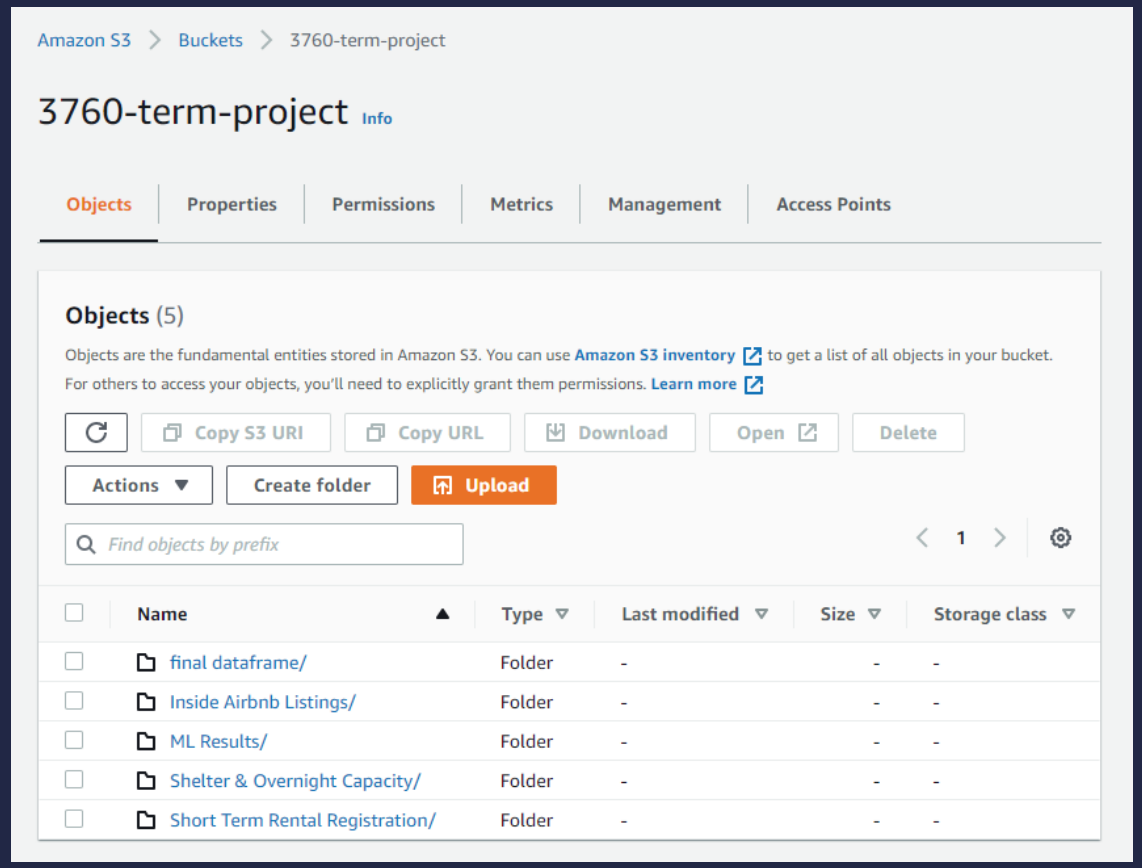
Both offers very similar capabilities

AWS Solution Architecture

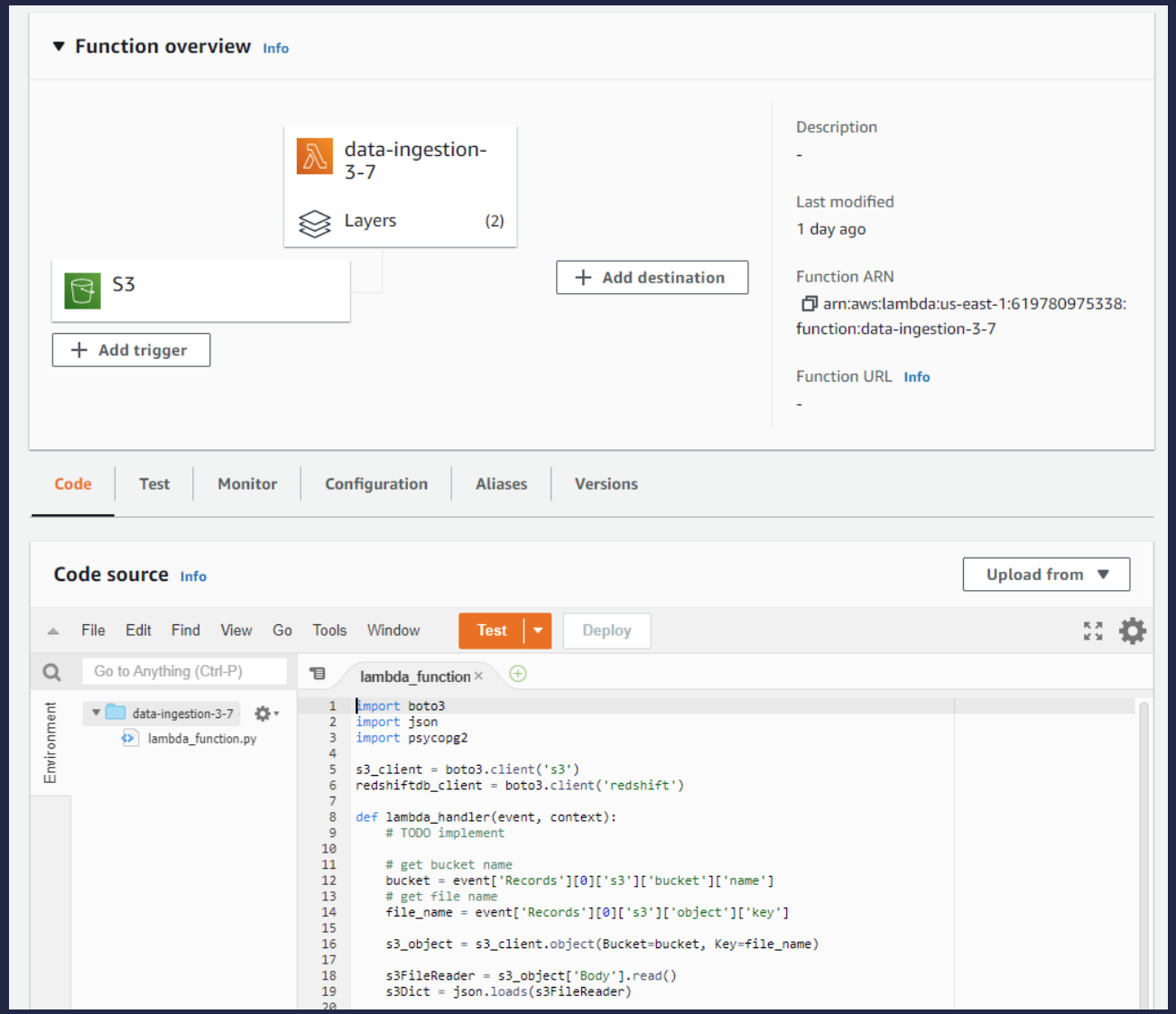


Implementation of Architecture

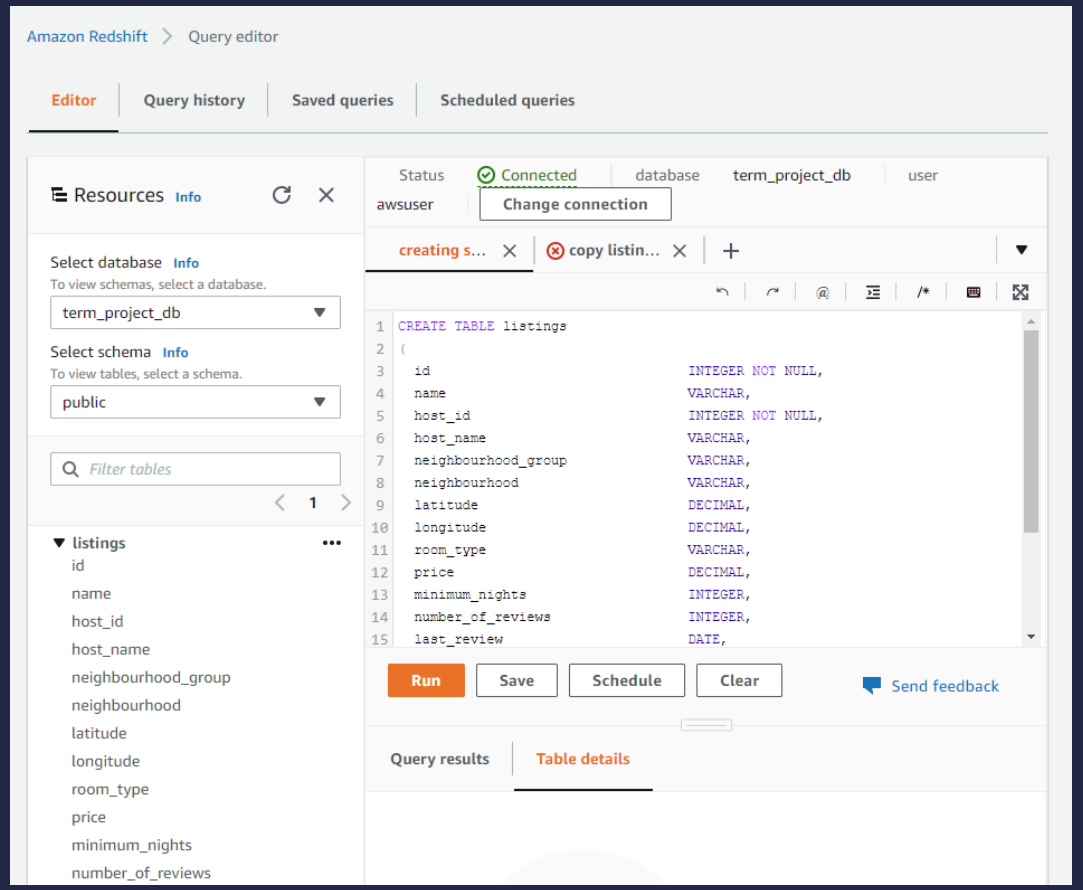
Amazon S3



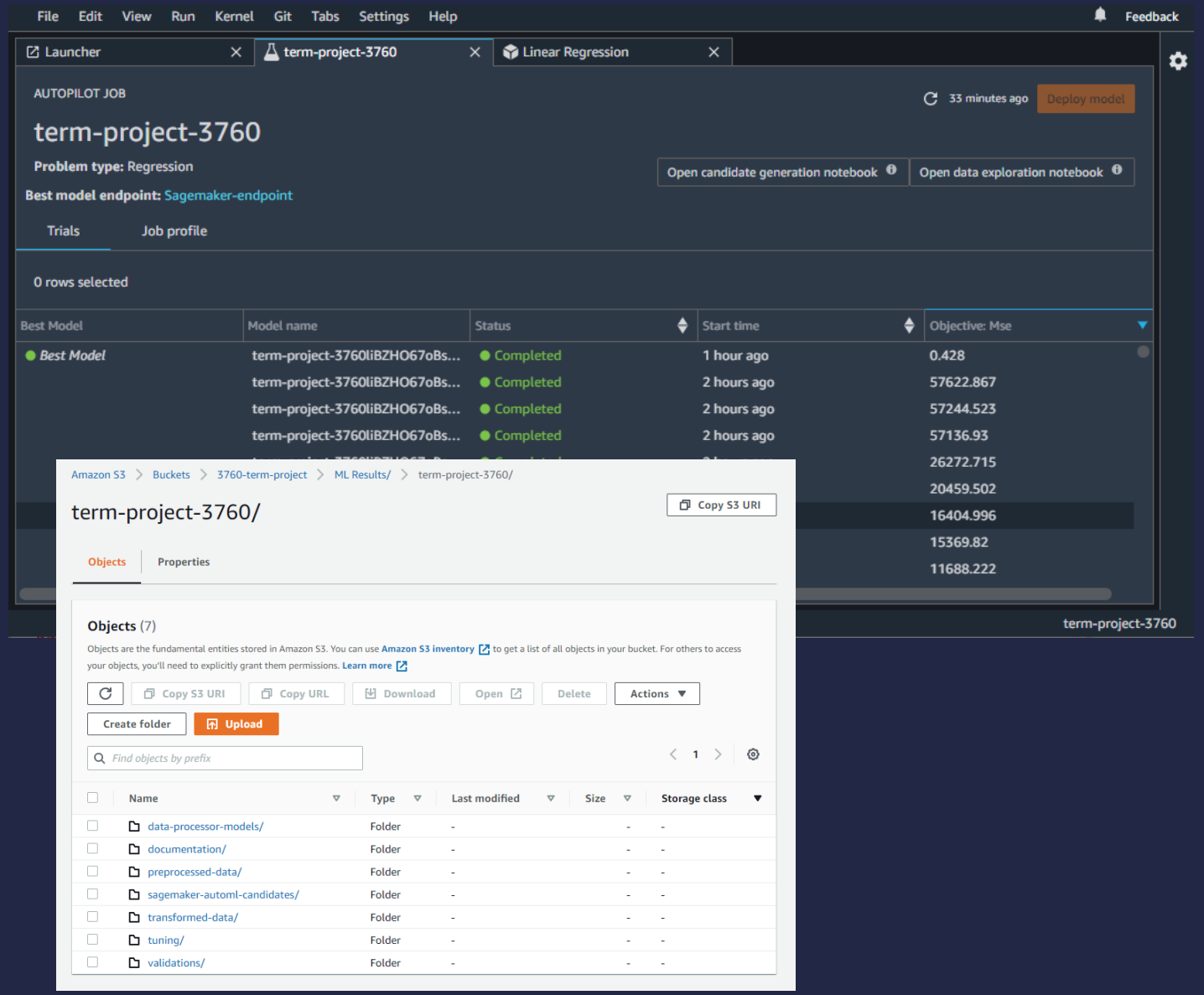
AWS Lambda



Amazon Redshift



Amazon SageMaker Autopilot



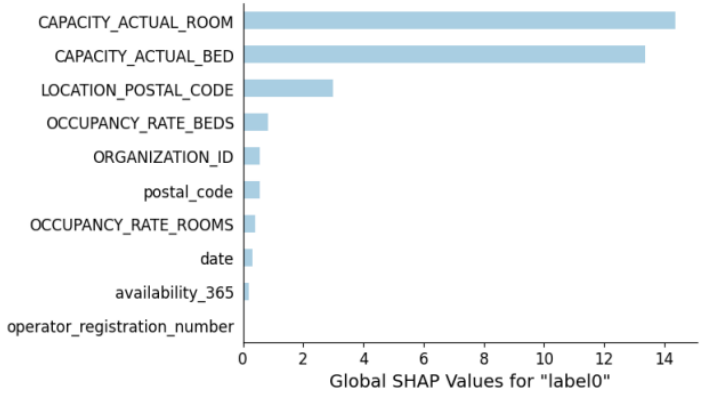
Amazon SageMaker Autopilot - Results

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.



jupyter SageMakerAutopilotDataExplorationNotebook (unsaved changes)

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted

Run Markdown

Suggested Action Items

- Investigate the origin of the data field. Are some values non-finite (e.g. infinity, nan)? Are they missing or is it an error in data input?
- Missing and extreme values may indicate a bug in the data collection process. Verify the numerical descriptions align with expectations. For example, use domain knowledge to check that the range of values for a feature meets with expectations.

	% of Numerical Values	Mean	Median	Min	Max	% of Outlier Values
SERVICE_USER_COUNT	100.0%	46.922	32.0	1.0	280.0	nan
ORGANIZATION_ID	100.0%	1.40851	1.0	1.0	7.0	3.2
OCCUPANCY_RATE_BEDS	69.33%	92.9049	98.0	9.09	100.0	nan
CAPACITY_ACTUAL_BED	69.33%	30.7609	25.0	2.0	234.0	nan
OCCUPANCY_RATE_ROOMS	33.8%	95.917	100.0	8.33	101.41	nan
CAPACITY_ACTUAL_ROOM	33.8%	60.4753	40.0	2.0	268.0	nan
operator_registration_number	30.38%	80.7411	49.0	1.0	451.0	nan
calculated_host_listings_count	30.38%	3.37813	3.05556	1.0	11.4828	nan
price	30.38%	162.601	149.633	45.0	950.0	nan
minimum_nights	30.38%	5.55265	4.78788	1.4	29.125	nan
reviews_per_month	30.38%	3.40869	3.22976	0.1	8.09333	nan
availability_365	30.38%	144.753	138.395	54.625	327.0	nan

Challenges

Challenges

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