DATS 6450 CLOUD COMPUTING

Airbnb Rental Price Prediction with AWS

Team

Aakash Hariharan
Neeraj Magadum
Trisha Singh

TABLE OF CONTENTS

01 Scope Of The Project

02 Features Implemented

03 Data Sources Used

04 Expected Outcomes

05 Architecture

06 References

SCOPE OF THE PROJECT

This project aims to develop a machine learning model using AWS SageMaker to analyze Airbnb listings and generate key insights. The model will predict:

OBJECTIVES

- Develop an automated pipeline for Airbnb price prediction.
- Identify key factors influencing listing prices.
- Deploy the model for real-time and batch predictions.

METHODOLOGY

- Store and preprocess data using Amazon S3 & AWS Glue.
- Train a model using AWS SageMaker.
- Deploy the model for predictions.
- Visualize insights using Amazon QuickSight.

TOOLS AND TECHNOLOGIES

- **AWS Services:** S3, Glue, SageMaker, QuickSight
- **Machine Learning**: XGBoost, Linear Learner

FEATURES IMPLEMENTED

STEP 1

Data Collection & Storage

The project will utilize **Amazon S3** as a central storage location for both raw and processed Airbnb data

STEP 2

Data Preprocessing & Feature Engineering

AWS Glue will be used to clean and transform the dataset before training. This includes removing inconsistencies, handling missing values, and standardizing features.

STEP 3

Model Training Using AWS SageMaker

The core of this project is training a regression model that can predict listing prices based on historical data.

STEP 4

Model Evaluation & Optimization

After training, the model's performance will be evaluated using Root Mean Squared Error (RMSE) and other regression metrics to assess accuracy.

STEP 5

Data Visualization & Insights

Once the predictions are generated, Amazon QuickSight will be used to create a dashboard and visualize the results.

DATA SOURCES

Dataset Overview

- Source : **Kaggle (**https://www.kaggle.com/datasets/arianazmoudeh/airbnbopendata/data)
- Contains 102,599 Airbnb listings with details on location, price, availability, and 23 such columns.
- The dataset also includes Airbnb listings from multiple neighborhoods and regions, allowing for location-based insights.
- There are some missing values and standardized text which needs to be pre-processed

Access In AWS

- In AWS Cloud, Amazon S3 storage service is used for it’s scalability and cost-efficiency.
- Directly integrates with AWS SageMaker for ML training.
- Supports QuickSight for BI visualization without needing databases.

EXPECTED OUTCOMES

01 Predicting fair and competitive prices can enhance customer satisfaction by balancing affordability and profitability.

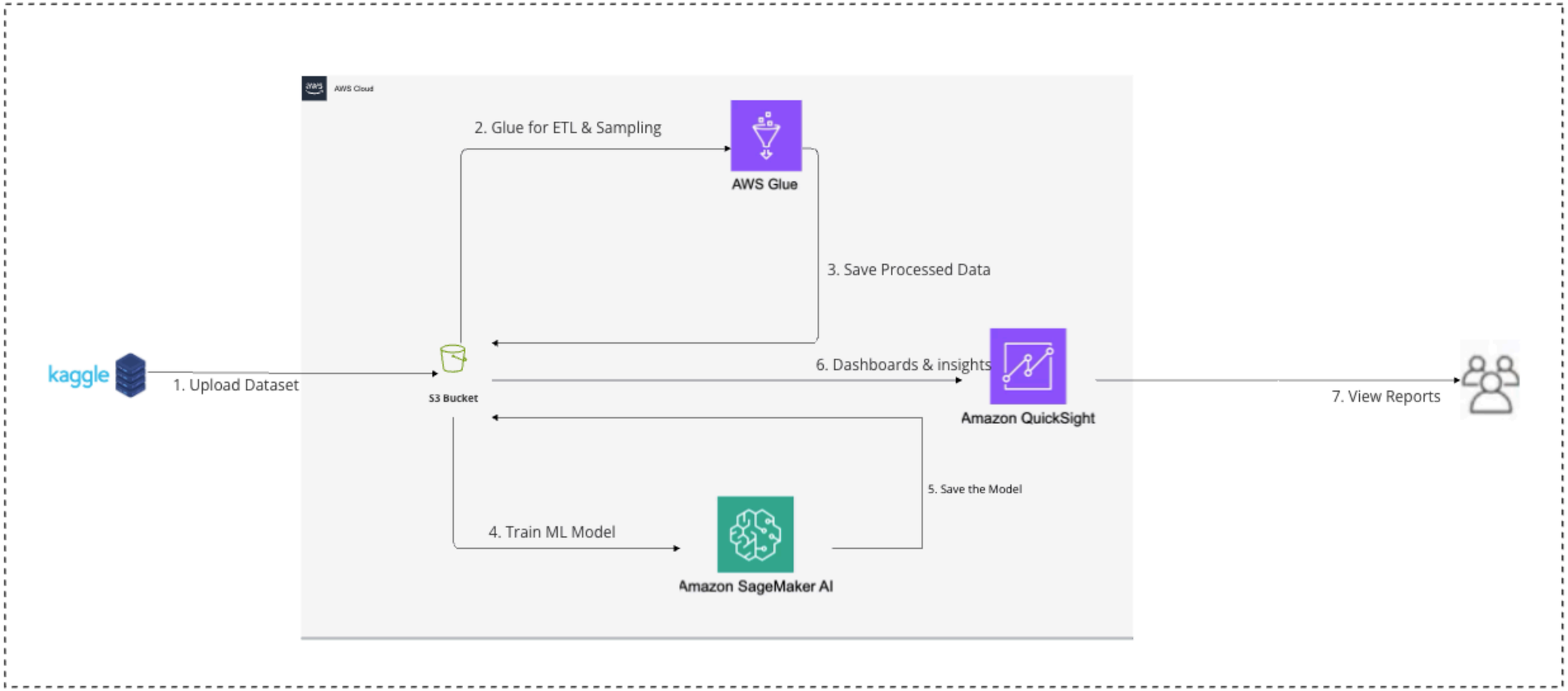
02 Insights into which variables (e.g., location, room type, availability) have the most significant impact on price.

03 A fully automated workflow using AWS services, ensuring efficient data processing, model training, and batch predictions.

04 Dashboards in Amazon QuickSight will display pricing trends

05 Using AWS services like SageMaker and Glue to ensure seamless processing and analysis of the Airbnb dataset.

ARCHITECTURE



PART II
Implementation

S3 Buckets

Image 1

Contains the project crawler

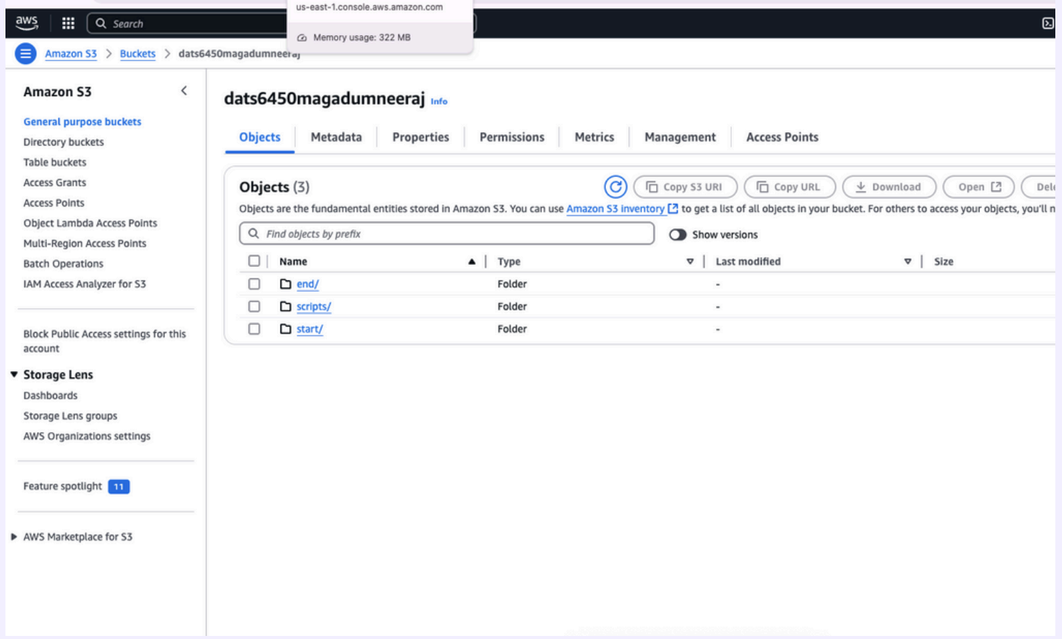
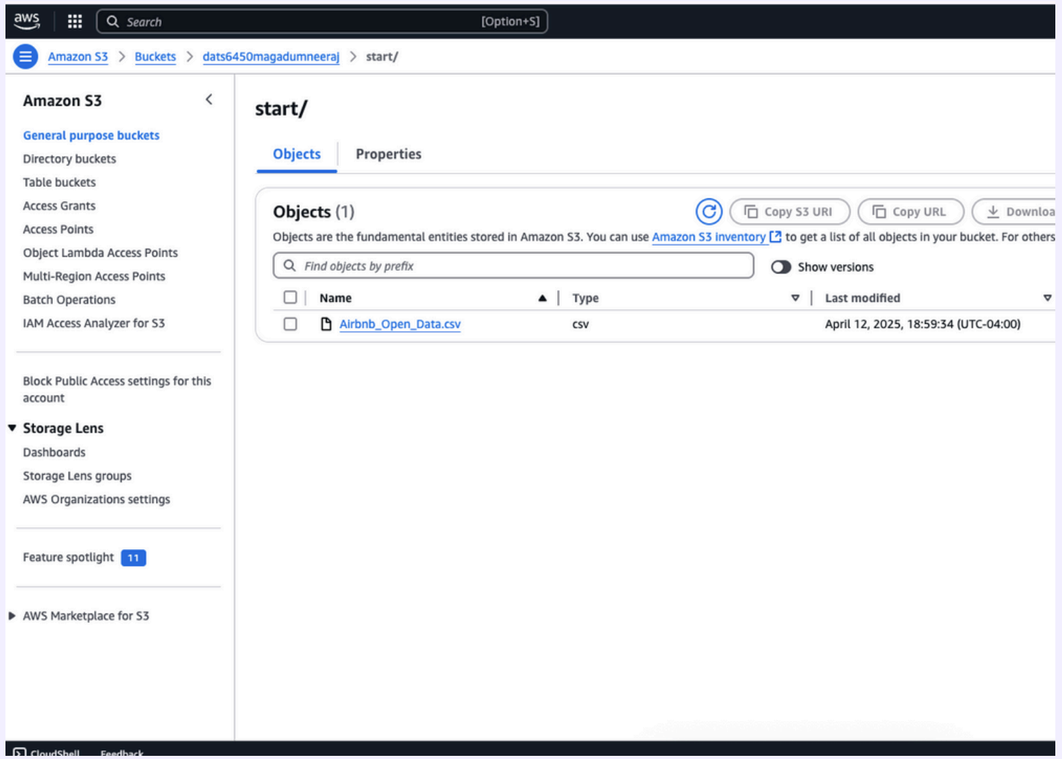
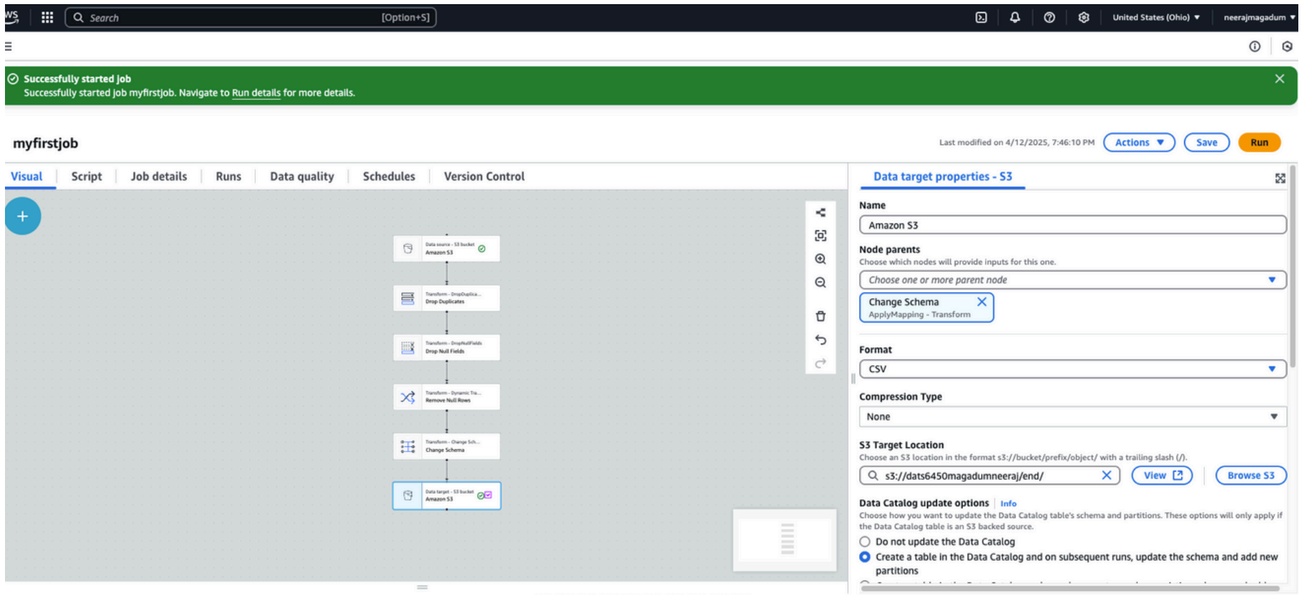


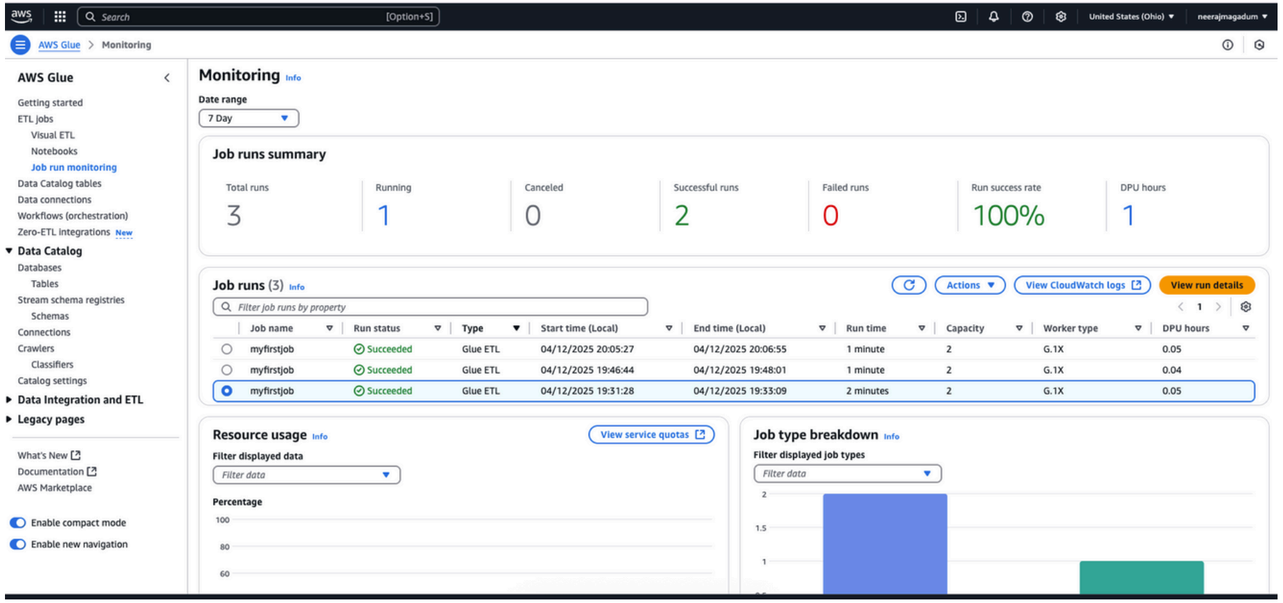
Image 2

Dataset stored in S3 Bucket

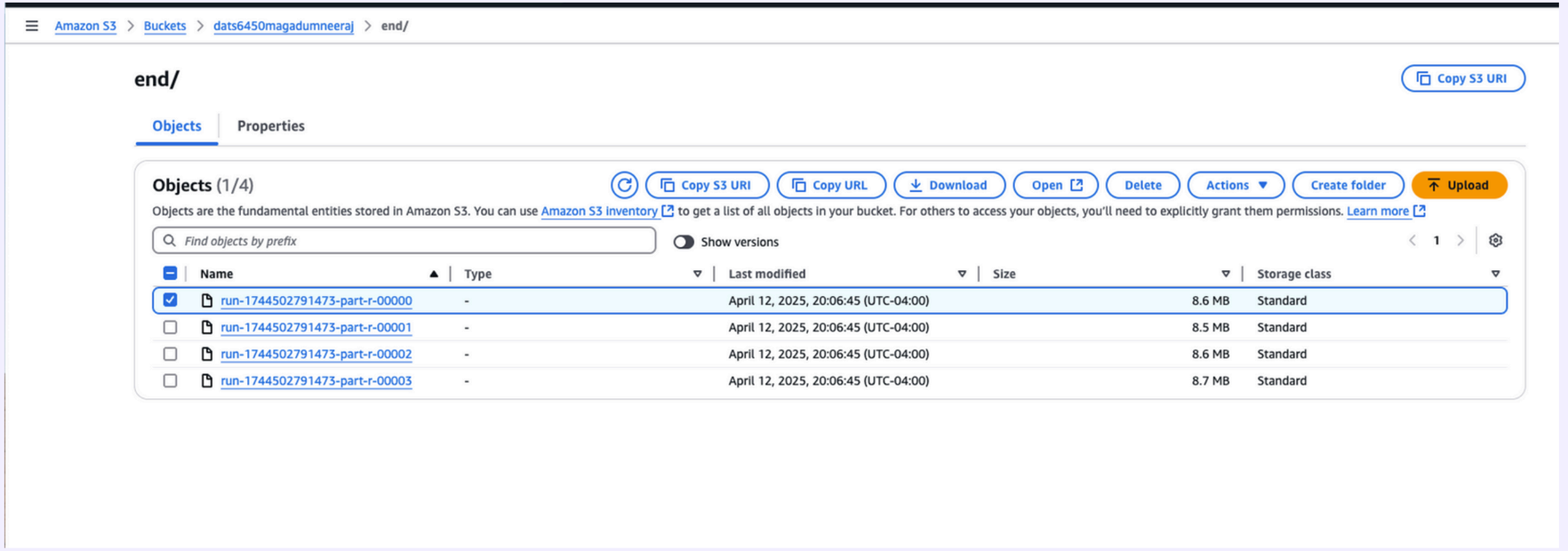


AWS Glue


Successfully started job
Successfully started job myfirstjob. Navigate to Run details for more details.
myfirstjob
Last modified on 4/12/2025, 7:46:10 PM
Actions
Save
Run
Visual
Script
Job details
Runs
Data quality
Schedules
Version Control
Data target properties - S3
Name
Amazon S3
Node parents
Choose one or more parent node
Change Schema
ApplyMapping - Transform
Format
CSV
Compression Type
None
S3 Target Location
s3://dats6450magadumneeraj/end/
View
Browse S3
Data Catalog update options
Do not update the Data Catalog
Create a table in the Data Catalog and on subsequent runs, update the schema and add new partitions

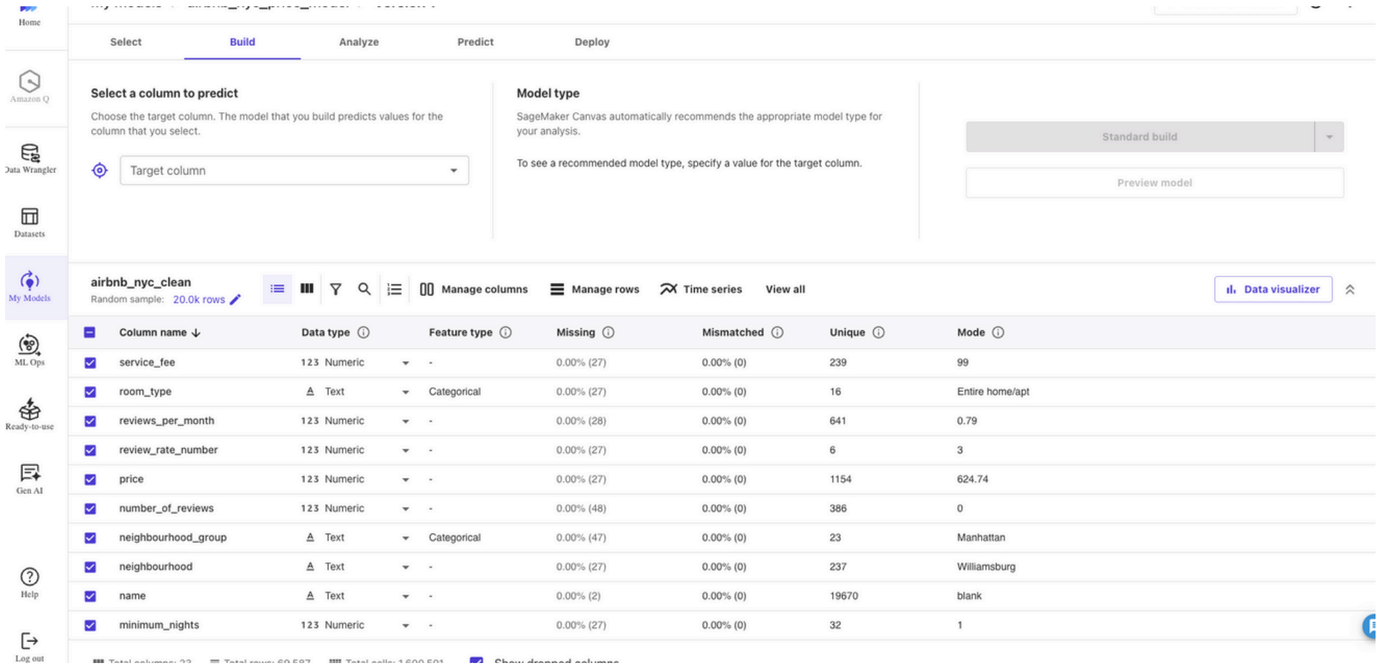

AWS Glue
Monitoring
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
Date range
7 Day
Job runs summary
Total runs 3
Running 1
Canceled 0
Successful runs 2
Failed runs 0
Run success rate 100%
DPU hours 1
Job runs (3)
Actions
View CloudWatch logs
View run details
Job name
Run status
Type
Start time (Local)
End time (Local)
Run time
Capacity
Worker type
DPU hours
myfirstjob Succeeded Glue ETL 04/12/2025 20:05:27 04/12/2025 20:06:55 1 minute 2 G.1X 0.05
myfirstjob Succeeded Glue ETL 04/12/2025 19:46:44 04/12/2025 19:48:01 1 minute 2 G.1X 0.04
myfirstjob Succeeded Glue ETL 04/12/2025 19:31:28 04/12/2025 19:33:09 2 minutes 2 G.1X 0.05
Resource usage
View service quotas
Job type breakdown

Results → S3


Amazon S3 > Buckets > dats6450magadumneeraj > end/
end/
Copy S3 URI
Objects
Properties
Objects (1/4)
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
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
Find objects by prefix
Show versions
1
Name
Type
Last modified
Size
Storage class
run-1744502791473-part-r-00000 - April 12, 2025, 20:06:45 (UTC-04:00) 8.6 MB Standard
run-1744502791473-part-r-00001 - April 12, 2025, 20:06:45 (UTC-04:00) 8.5 MB Standard
run-1744502791473-part-r-00002 - April 12, 2025, 20:06:45 (UTC-04:00) 8.6 MB Standard
run-1744502791473-part-r-00003 - April 12, 2025, 20:06:45 (UTC-04:00) 8.7 MB Standard

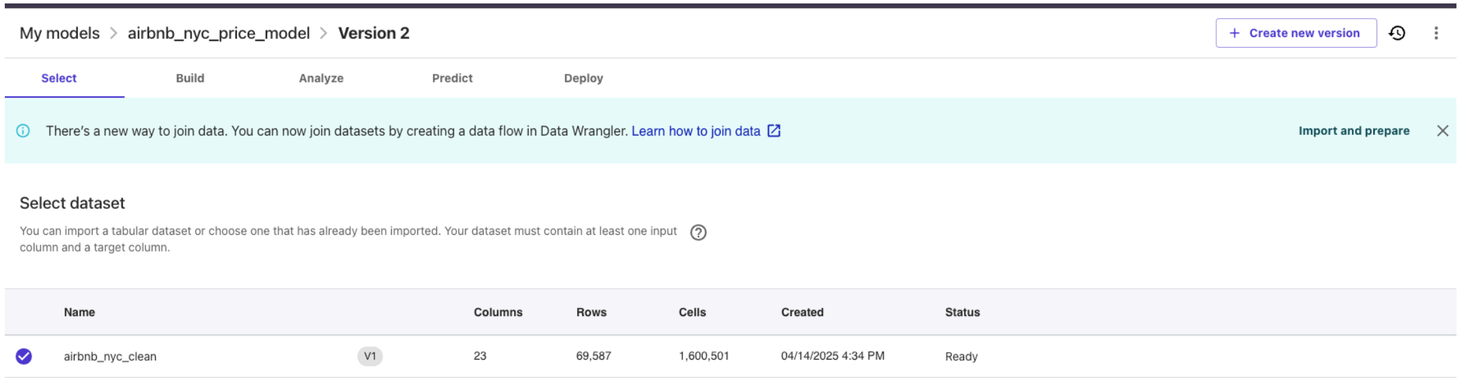
AWS Sagemaker AI - Canvas

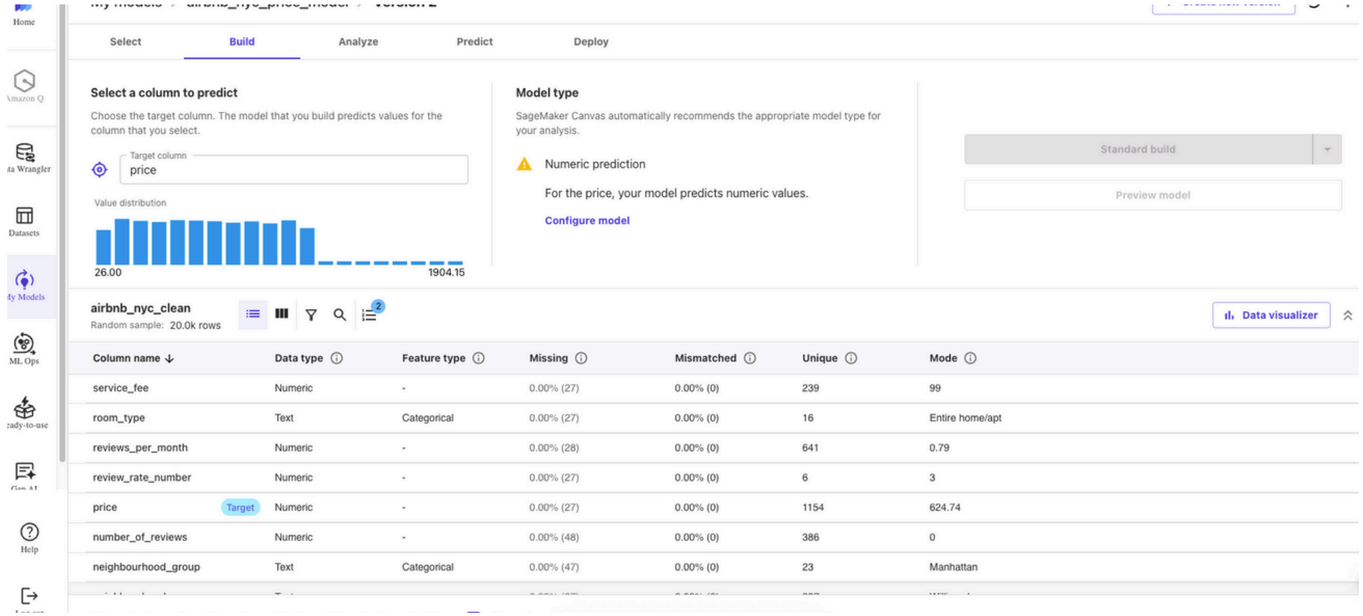
Add more text



Reslts of Model Trained

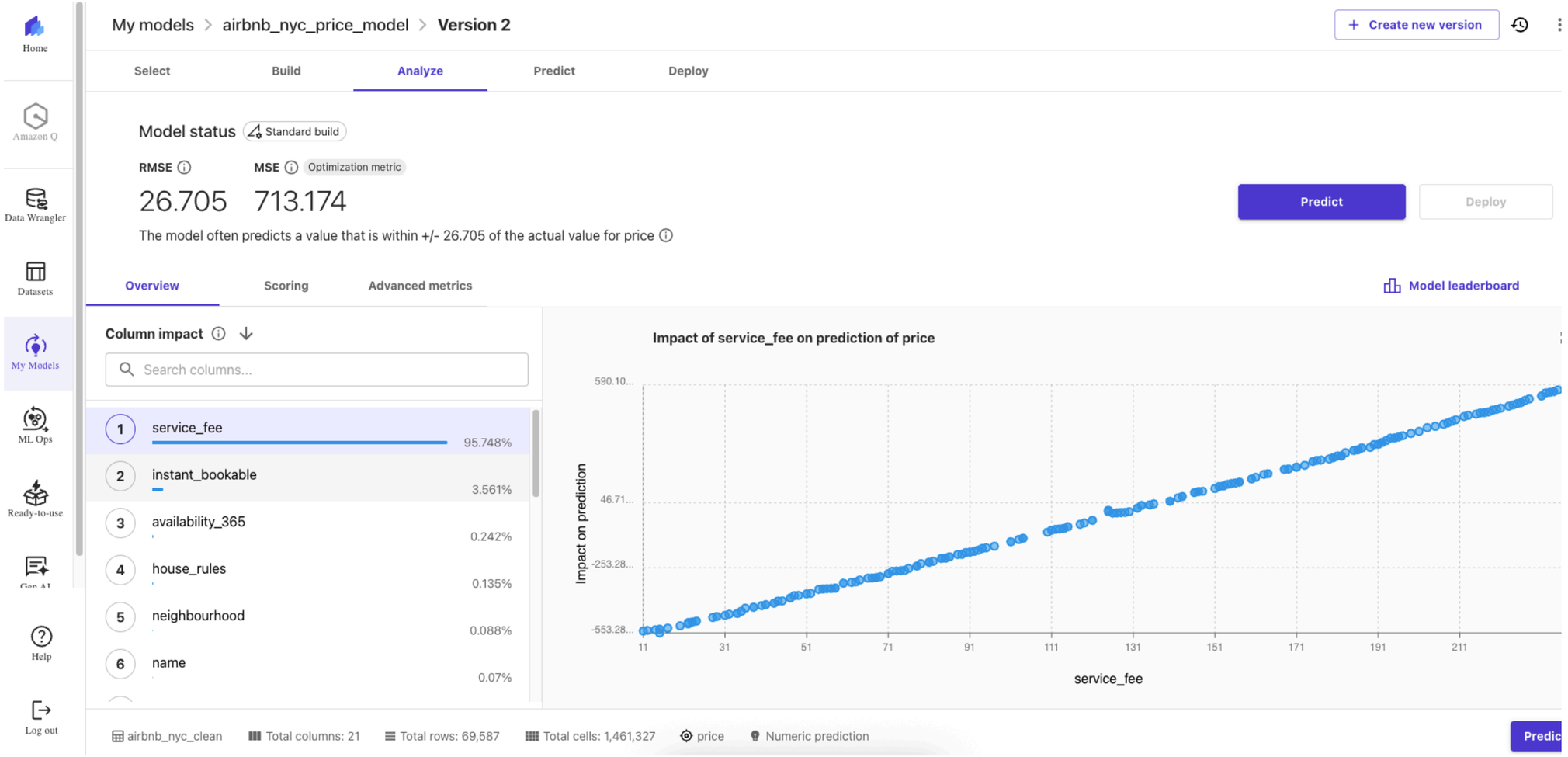
Select Model



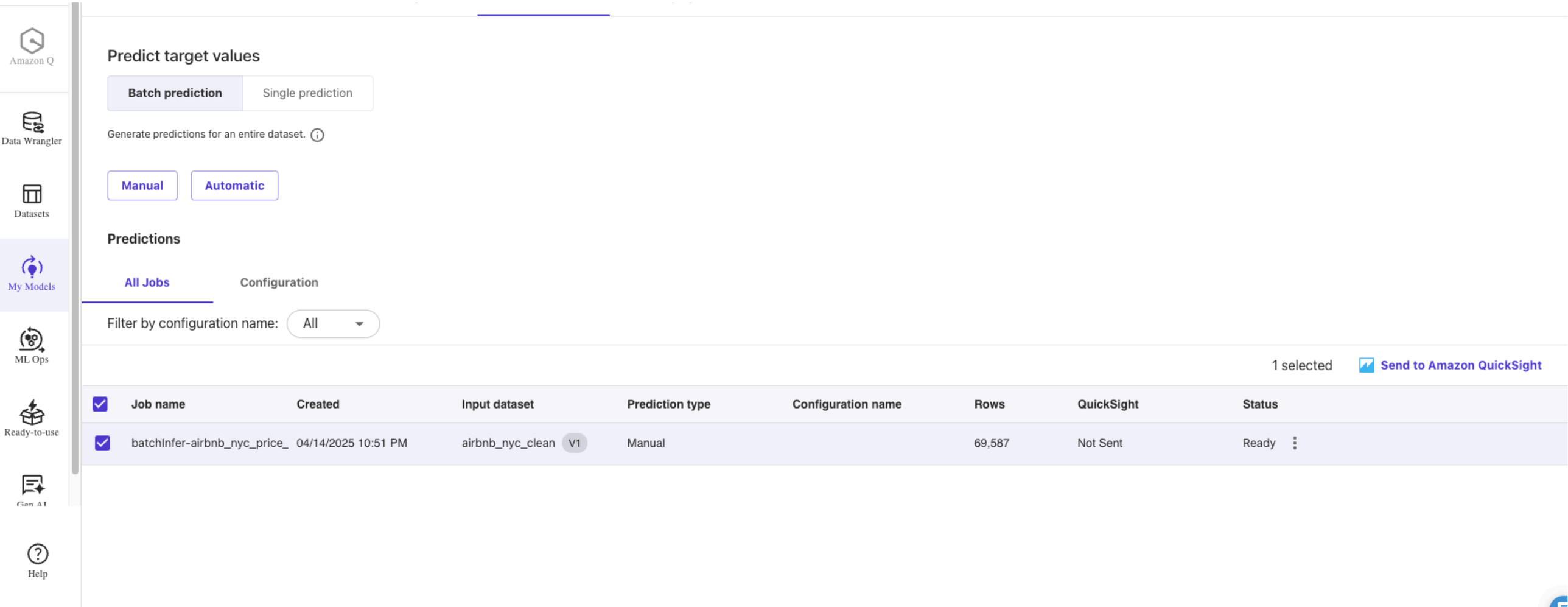

Home
Amazon Q
Data Wrangler
Datasets
My Models
ML Ops
Ready-to-use
Help
Select
Build
Analyze
Predict
Deploy
Select a column to predict
Choose the target column. The model that you build predicts values for the column that you select.
Target column
price
Value distribution
26.00
1904.15
Model type
SageMaker Canvas automatically recommends the appropriate model type for your analysis.
Numeric prediction
For the price, your model predicts numeric values.
Configure model
Standard build
Preview model
airbnb_nyc_clean
Random sample: 20.0k rows
Data visualizer
Column name
Data type
Feature type
Missing
Mismatched
Unique
Mode
service_fee Numeric - 0.00% (27) 0.00% (0) 239 99
room_type Text Categorical 0.00% (27) 0.00% (0) 16 Entire home/apt
reviews_per_month Numeric - 0.00% (28) 0.00% (0) 641 0.79
review_rate_number Numeric - 0.00% (27) 0.00% (0) 6 3
price Target Numeric - 0.00% (27) 0.00% (0) 1154 624.74
number_of_reviews Numeric - 0.00% (48) 0.00% (0) 386 0
neighbourhood_group Text Categorical 0.00% (47) 0.00% (0) 23 Manhattan

Build Model

Model Evaluation & Insights

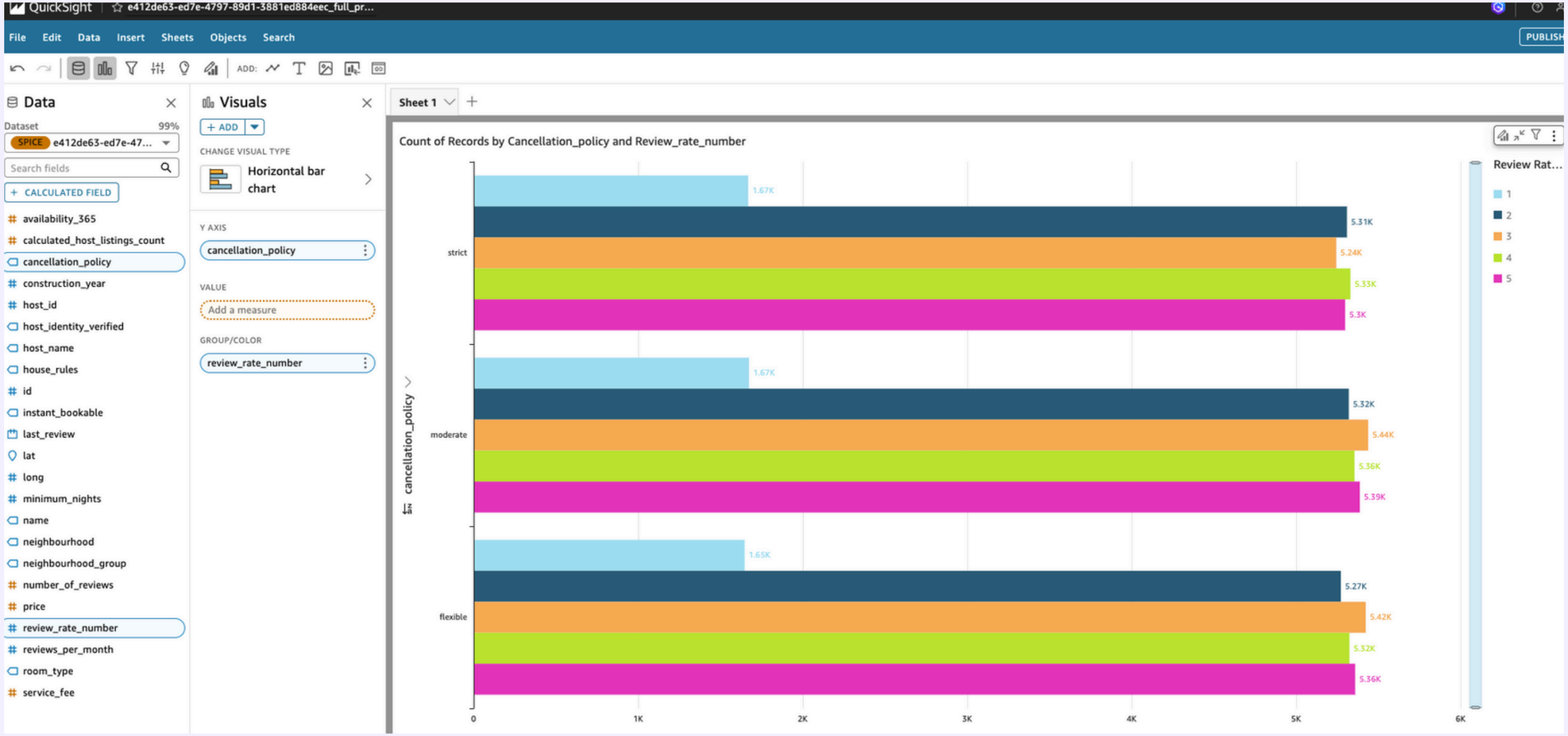


Predict Model

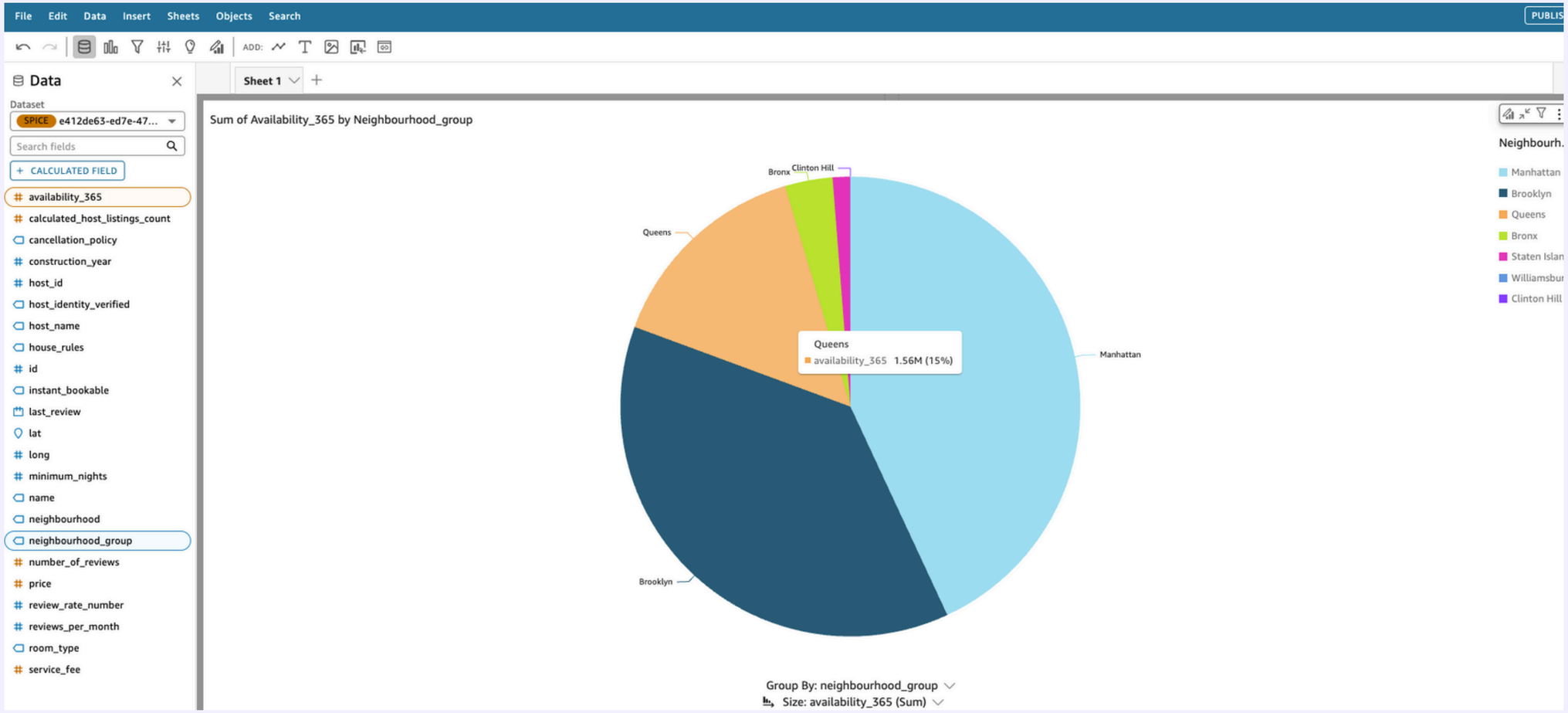


AWS Quicksight Visualizations

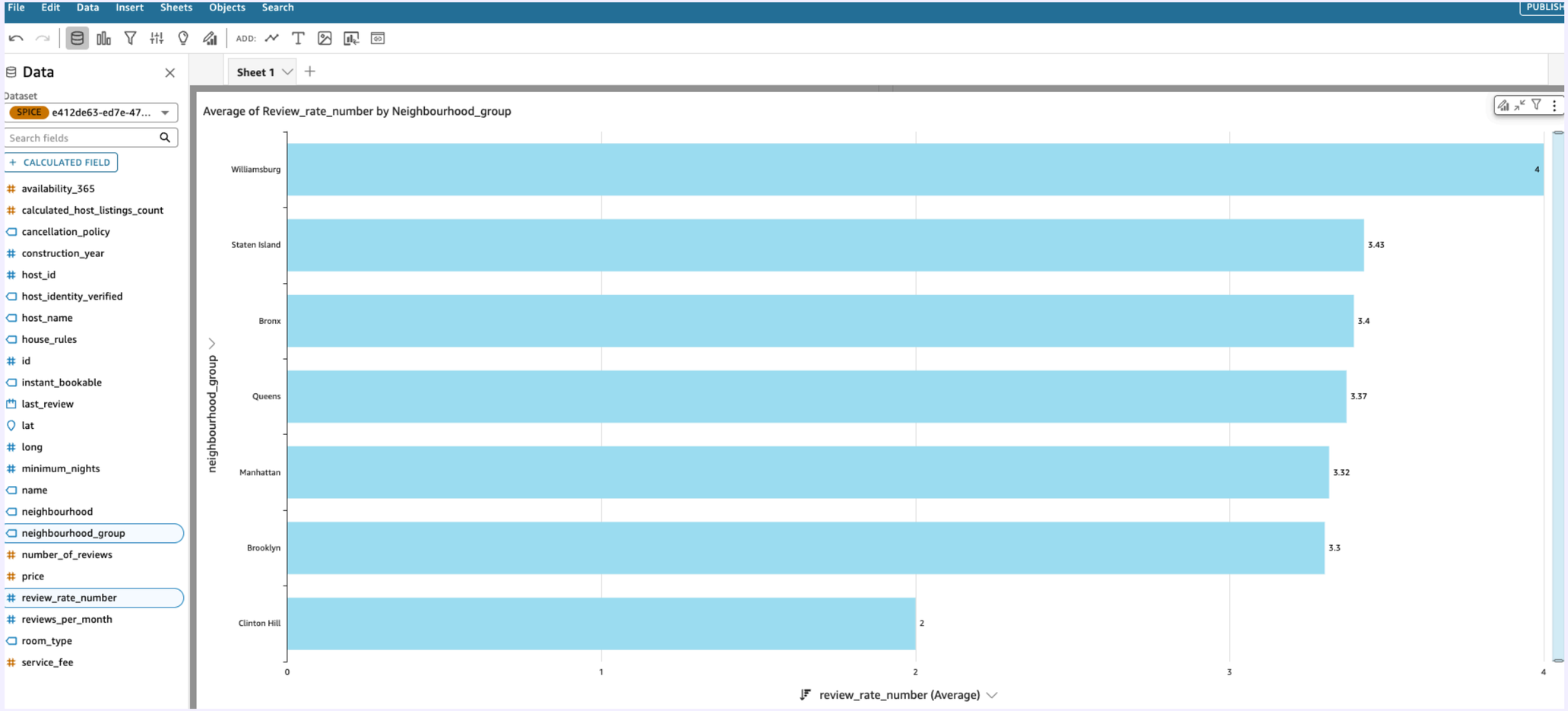
Impact of Cancellation Policies on Review Ratings



Distribution of Property Availability by Neighborhood Group



Average Review Rating by Neighborhood Group



REFERENCES

01	https://www.kaggle.com/datasets/arianazmoudeh/airbnbopendata/code
02	Prajapati Pradip, & Prof. Monali Suthar. (2022). A Survey On Price Prediction Model for Airbnb listing using Machine Learning. International Journal of Scientific Research in Science, Engineering and Technology, 167–171.
03	Case Study of Airbnb using AWS Cloud. (n.d.). Www.linkedin.com. https://www.linkedin.com/pulse/case-study-airbnb-using-aws-cloud-prashant-singh/
04	AWS. (2019). Airbnb Case Study – Amazon Web Services (AWS). Amazon Web Services, Inc. https://aws.amazon.com/solutions/case-studies/airbnb-case-study/
05	Amazon. (2024). ML on AWS – Maximize Outcomes with Machine Learning and AI – AWS. Amazon Web Services, Inc. https://aws.amazon.com/ai/machine-learning/
06	Pouya Rezazadeh Kalehbasti, Liubov Nikolenko, & Rezaei, H. (2019, July 29). Airbnb Price Prediction Using Machine Learning and Sentiment Analysis. https://doi.org/10.48550/arXiv.1907.12665

GITHUB REPO

https://github.com/Aakash2112/Airbnb-Rental-Price-Prediction-with-AWS.git