Group 7- Team Members

Ashmitha Paruchuri Balaji

Nikitha Seelam Balaji

Shashira Guntuka

Shreya Akotiya

PROJECT PROPOSAL

Improving Urban Livability: An Analysis of 311 Complaints and Their Relationship to Short-Term Rentals

ABSTRACT

The aim of this project is to build a system that looks at 311 complaints across the cities of the United States and explores how short-term rentals like Airbnb, along with tourism in general, may be linked to these complaints. The idea is to understand if there are patterns between rental activity and issues that affect day-to-day life such as noise, sanitation problems or illegal occupancy. To achieve the project will combine two main datasets: 311 service requests and Airbnb listings/booking data. By combining these dataset we aim to determine whether more 311 complaints are also reported in areas with higher short-term rental activity. The results gives us useful insights for ci, local community groups, and planners who are trying to make the most of tourism while still keeping neighborhoods comfortable livable for residents.

MOTIVATION

Across many cities in the United States the rise of short term rentals such as Airbnb has created both opportunities and challenges. These platforms increase the flexibility of tourism, improve the economy and provide new ways of income to the property owners. Also the short-term rentals can make life difficult in neighborhoods where residents are already dealing with everyday concerns. Problems like noise from house parties or construction activities, sanitation issues, parking shortages, and even housing affordability are becoming more common, and they are often reported through 311 complaint systems.

New York City is one of the appropriate cases where these challenges can be seen. Tens of thousands of Airbnb postings have shown a noticeable overlap between neighbourhoods that have significant rental activity and those with higher resident complaint volumes. But this isn't

just a New York problem; similar patterns are showing up in cities like San Francisco, Los Angeles, and Chicago, where residents feel the strain of tourism on local services.
The motivation behind this project is to understand these connections in a systematic way by combining complaint data with short term rental data. By doing so, we can highlight how tourism and short term rental activity affect urban livability across U.S. cities and give policymakers and communities better evidence to design solutions that balance economic growth with residents' quality of life.

LITERATURE SURVEY

- Lee, S., & Kim, H. (2023). Four shades of Airbnb and its impact on locals: A spatiotemporal analysis of Airbnb, rent, housing prices, and gentrification. Tourism Management Perspectives, 49, 101192: Lee and Kim analyze how Airbnb listings influence housing affordability and neighborhood change. Their study shows how entire home rentals and professional operators push up rents and home prices most sharply, while smaller hosts have only minor effects.
  Relevance: This work shows how short-term rentals do not affect all cities in the same way, providing evidence that regulations should focus on high-impact operators rather than casual hosts.
  Link : https://www.sciencedirect.com/science/article/abs/pii/S2211973623001204
- "Detailed Data Analysis: The Rise of NYC 311 Noise Complaints" (NYC Data Science blog):This study explains the trends in noise related 311 calls in New York City, understanding the temporal patterns, spatial hotspots, and potential correlates .The analysis also tells us that noise complaints have grown in certain towns and cluster in neighborhoods with higher density or lower incomes.
  Relevance: This research provides insights into how open city data can be used to identify urban quality-of-life concerns, offering valuable information about the relationship between neighborhood characteristics and resident well-being.
  Link:https://nycdatascience.com/blog/student-works/detailed-data-analysis-the-rise-of-nyc-311-noise-complaints/
- **"**Equity in 311 Reporting: Understanding Socio-Spatial Differentials in the Propensity to Complain" (Kontokosta, Hong, & Korsberg, 2017, arXiv): This paper takes a detailed look on how New York City residents use the 311 system to report problems .The authors argue that complaint data, often used by researchers and policymakers as a window into neighbourhood conditions does not always reflect actual living conditions because not everyone who report issue are at the same rate. They even mention that under-reporting is more common in lower income regions , minority, and limited-English-speaking areas, while over-reporting tends to occur in wealthier, whiter, and more educated

neighbourhoods. The study demonstrates that complaint records capture both the presence of problems and the social dynamics of who feels able or empowered to speak up.
Relevance: For projects that rely on 311 complaints such as analysing the neighbourhood effects of short-term rentals this work offers an important understanding as it shows that raw complaint counts can be misleading unless we account for differences in reporting behaviour such as under reporting and it provides a framework for doing the same .
Link:https://arxiv.org/pdf/1710.02452

- The perceived impacts of short-term rental platforms: Comparing the social exchange perspective" (2024, International Journal of Sociology and Social Policy):This article tells us how residents view the growth of short-term rental platforms like Airbnb in their neighborhoods. The authors also use the social exchange framework to compare perceived benefits such as economic opportunities and tourism income against drawbacks that it can bring such as noise, loss of community cohesion or higher rental/housing costs. The findings highlight that perceptions vary across different resident groups consisting both social and economic divides.
  Relevance: This study help us gain useful insight into how communities balance the positives and negatives of short-term rentals showing people's opinions about Airbnb are mixed and depend a lot on where they live. The study shows that Airbnb isn't only about money and tourism etc but also about how it changes everyday community.life."Link:https://www.sciencedirect.com/science/article/pii/S0160791X24001349?utm_source=chatgpt.com
- "The relationship of Airbnb to neighborhood calls for service in three cities" (2021, Cities): This study examines whether the presence of Airbnb rentals is related with changes in neighborhood calls for service including the 311 complaints and police responses in three U.S. cities (Nashville, New Orleans, and Portland). Using spatial analysis, the authors shows that higher concentrations of Airbnb listings are linked to increases in certain types of service calls, particularly those related to public disturbance and disorder. The study suggests that short-term rentals may put added pressure on local governments and services.
  Relevance: This research directly connects Airbnb activity to urban service data, making it highly relevant for projects studying 311 or 911 complaints. It shows that the growth of short-term rentals can translate into measurable impacts on neighborhood livability and city resources.
  Link:https://www.sciencedirect.com/science/article/abs/pii/S0264275121001414?utm_source=chatgpt.com

---

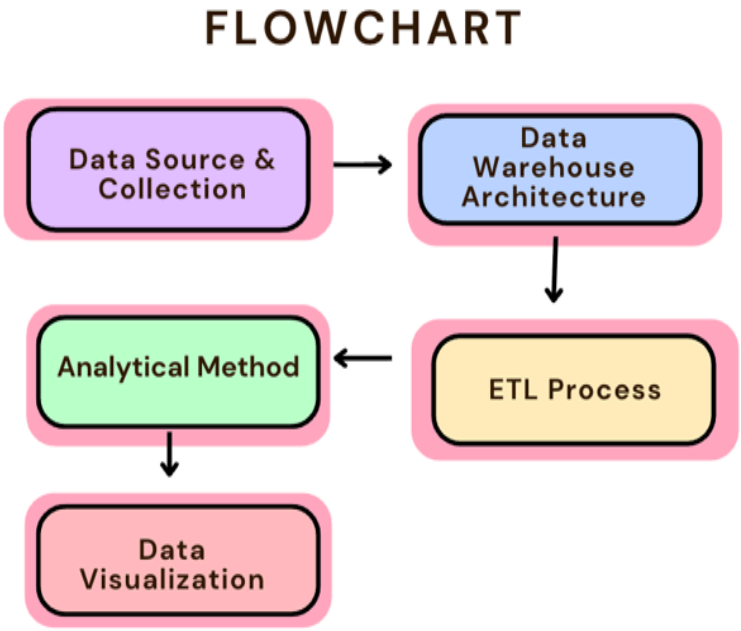
METHODOLOGY

The project's main goal is detailed analysis of 311 city complaints in New York City to understand the factors that contribute to these complaints. Secondly we will integrate Airbnb listings to explore the correlation between short-term rental activity and specific types of complaints, such as noise , sanitation, parking etc.

Our approach is structured in two main phases:

Phase 1: Analysis of 311 Complaint Data, The first step is primary analysis of the 311 data by identifying complaint hotspots, determining the most common types of complaints reported, and understanding trends. We will use more additional data to understand complaint activity.

Phase 2: Data Integration and Correlation , we will integrate the 311 data publicly available with Airbnb listings and will do comparative analysis of **New York City 311 complaint data** and **Airbnb listings data** to identify patterns, gaps, and challenges. This enrichment with airbnb data will help us to see if 311 complaints areas vary with airbnb activity , meaning high density areas with higher volume of complaints. By using statistical methods, we will explore potential correlations, ultimately providing actionable insights for city policymakers to improve urban livability.



Data Sources

We will collect data from trusted sources, including **NYC 311 complaints** and **InsideAirbnb** listings and calendars.

- **NYC 311 complaints** (NYC Open Data)

  This dataset gives all the service-related complaints in New York city. It includes **t**imestamps, complaint types (e.g., noise, sanitation, illegal rentals), location and resolution status.

- **InsideAirbnb** listings and calendars (insideairbnb.com)

  This dataset will help to measure the density and activity level of short-term rentals in New York city. It includes Airbnb properties with details like location, room type, price , number of reviews , daily availability , pricing etc.

Using ETL processes, we'll organize the data for trend analysis and finding potential causal relationships between the two datasets by aggregating datasets by locations to compare STR activity and complaint volume over time.

Data Warehouse Architecture

1.**Fact Table : Neighborhood_Pressure_Fact**
This is the fact table, and it contains all the measures such as weekly/monthly complaint counts**,** STR counts**,** occupancy rate**,** average daily price**,** and the Neighborhood Pressure Index (NPI)**.** The fact table is aggregated at **(**time_id, location_id)—choose week or month for trend comparisons and analysis.

2. **Dimension Tables :** These tables contain all the descriptive data regarding attributes.

- **Location_Dimension:** Geographic attributes (neighborhood, borough, Zipcode).
- **Complaint_Dimension:** Complaint types (type, subtype, severity, responsible agency).
- **STR_Dimension:** Listing characteristics (property type, room type, host category/status).
- **Agency_Dimension:** Information about city agencies responsible for addressing complaints, including their code (e.g., NYPD) and full name.
- **Provider_Dimension:** Platform information (Airbnb).
- **Listing_Dimension :** Listing attributes (capacity, amenities, license flag).
- **Host_Dimension :** Host attributes (superhost flag, tenure bucket).

ETL Process

For this project, we set up an ETL pipeline that collects together data from two different sources: InsideAirbnb listings and 311 service request records for New York location.

- **Extract:** Data is ingested directly from public APIs and CSV sources in raw form
- **Transform:** In this step, the data is cleaned and reformatted. Will make dates aligned to weeks or months, addresses will be matched with neighborhood IDs etc.
- **Load:** The processed data is loaded into a star schema/ or snowflake . We use incremental updates so that only new or changed records are added, which makes the pipeline faster and keeps a history for time-based analysis.

Analytical Methods

Will analyze how city service complaints relate to short-term rental activity by looking into measures like the number of active listings, occupancy rates and average prices, compare them with complaint types , count etc. By looking at these metrics over time and across neighborhoods, our aim is to highlight patterns and explore possible connections between the two datasets.

Data Visualization

The result will be presented through interactive dashboards/ stories. These will include time-series plots,bar charts, geographic visualizations of outcome variations ,neighborhood-level comparisons to show how complaints and STR activity vary. The goal is to make it easy to see trends, spot differences across areas, and better understand how these two factors interact.

Tools & Technologies

- **Programming and Data Processing:** Python for cleaning and transformation , Alteryx for merging.
- **Storage:** Amazon S3 buckets storage
- **Data Warehouse:** Amazon Redshift or Snowflake
- **Workflow**: Dagster/Apache Airflow ETL workflows
- **NoSQL Layer:** MongoDB, Neo4j .
- **Visualization**: Dashboards using Python (matplotlib, seaborn) and PowerBI for interactive exploration.

DELIVERABLES

Milestone	Deliverables
Week 1 & Week 2	· Problem Definition and Dataset Selection · Initial Project Proposal · Data ingestion setup using **Python APIs**
Week 3 & Week 4	· **ETL Pipeline with Python and Altyrx** · Data Cleaning and Transformation · Workflow orchestration using **Dagster / Prefect**
Week 5 & Week 6	· Data warehouse on **Amazon Redshift** · Storage integration via S3 · Scheme design -> **Star/Snowflake Schema** · NoSQL layer with **MongoDB, Neo4j**
Week 7 & Week 8	· Developing Dashboards using **Python & PowerBI** · Key spatial - temporal insights · Draft report & presentation
Week 9	· Final presentation & Project Elevator Pitch

TEAM MEMBERS AND ROLES

Team Member	Roles
Ashmitha Paruchuri Balaji	Data Pipeline Design, Orchestration(Dagster/Prefect)
Nikitha Seelam Balaji	Data Analysis, Feature Engineering, Insight Generation
Shashira Guntuka	Data Warehousing(Amazon Redshift), Schema Modelling, Integration
Shreya Akotiya	Data Collection, ETL Implementation & Optimization, Query Performance

RELEVANCE TO THE COURSE

This project applies core data warehousing and ETL concepts to a real-world urban policy problem. We use fact and dimension tables, star schema design, and slowly changing dimensions to organize short-term rental and 311 complaint data for effective analysis. The ETL process demonstrates how to extract, clean, and transform large datasets, which includes handling missing values, inconsistent formats, and spatial-temporal joins.

The project shows how both relational and NoSQL systems can work together, by integrating Amazon Redshift, S3, and MongoDB. We will apply query optimization and data quality checks. Finally, the project will show how data warehousing and visualization tools can produce meaningful insights for urban livability and policy making, linking technical concepts directly to real world impact.

TECHNICAL DIFFICULTY

The main challenges in this project come from working with large and messy datasets. Since both short term rental data and 311 complaints come from different sources, one major difficulty would be data cleaning which includes handling missing values, inconsistent formats, etc. Another challenge would be data integration, where 311 complaints will be in the form of text which is unstructured data and structured data would be in the form of tables. These both different types of data must be combined into a single warehouse.

Building a robust ETL pipeline is also complex, as it requires time based aggregations and transformations efficiently. As the data is high volume and continuously updated we may also expect performance issues. To handle this, techniques like indexing, partitioning, and parallel processing will be needed.

NOVELTY

1. Comprehensive Data Integration Strategy

First structured integration of NYC 311 complaints data and InsideAirbnb listings data at the neighborhood level of New York city. Development of a Neighborhood Pressure Index (NPI) that quantifies the correlation between short-term rental density and urban service complaints. Application of temporal analysis across multiple years to identify seasonal trends and long-term tendencies.

2. Multi-dimensional Correlation Analysis

Goes beyond simple correlation to examine causal relationships between Short-term rental (STR) activity (occupancy rates, prices, listing density) and complaint categories of interest. Incorporates host characteristics (superhost status, tenure) and property characteristics (capacity, amenities) as moderating variables Uses advanced statistical methods to control for confounding variables like neighborhood characteristics and seasonal trends.

IMPACT

The diverse implications of a data-informed approach to drive policy on city short-term rentals. The analytical framework is built with the intention of **assisting city policymakers** by providing evidence for **focused intervention, allocation of resources, and zoning in affected areas**. Scholarly, the approach offers an actionable template of **livability factor analysis** and adds to sharing economic research on its **implications for New York city**. Besides, the results promote openness and advocacy through enabling residents and grassroots organizations to harness data when raising **residents' concerns of quality of life to policymakers**. Lastly, the model aims at achieving a **balanced regulatory** scenario that **promotes tourism** as well as **preserves housing supply and promotes sustainable urbanization**.

THE HEILMEIER CATECHISM

1. **What are you trying to do in this project? Articulate your objective.**

   We aim to quantify and understand the **connection** between **short-term rentals** and **how liveable urban areas truly are** by looking at NYC 311 complaints alongside Airbnb activity. We are trying to find out if neighborhoods that have a lot of short-term rentals (STR) have more complaints about their quality of life. Plus, we're keen to pinpoint which types of complaints are most closely linked to Short-term rental activity.

2. **How is it done today? What are the current day limitations?**

   Right now, urban planners and policymakers depend on informal data, isolated analyses of individual datasets or limited surveys. Most research tends to look at 311 complaints or short-term rental activity separately. When they do explore correlations, they usually stick to basic statistical methods without considering important factors like time, location, or demographics.

3. **What are your focus points in your findings?**

   Our approach uniquely integrates two extensive and completely different datasets with analytical measures:

   (a) systematic merging of various large datasets

   (b) the creation of a new Neighborhood Pressure Index

   (c) a multi-faceted correlation analysis that includes host and property details, and

   (d) a scalable ETL architecture specifically crafted for urban policy research.

4. **Who cares and how is it useful to them?**

   This Project helps **Policy makers** who are struggling with issues ranging from housing shortages, the downside of mass tourism, and resentment from long-term residents of being displaced. **Urban planners** dealing with infrastructure, resolving noise problems, and attempting to ensure that emergency services can manage in tourist hotspots. **Locals and community groups** are meanwhile suffering a decline in living standards due to short-term neighbors and party houses. **Housing campaigners** are fighting hard against gentrification and maintaining a stock of low-cost housing. **Tourism authorities** are trying to strike a balance between economic gain from tourism and the need for sustainable visitor dispersal. **Emergency services** trying to decide how to allocate resources in areas where the population can shift suddenly. **Scholarly academics** are also

exploring research on urban economics and the influence of the sharing economy. San Francisco, Amsterdam, and Barcelona are also implementing similar short-term rental regulations.

5. **What are the risks that can be encountered in the project?**

   Data quality issues: The 311 complaints data has historic data from 2010 which might have skewed or partial reporting/missing complaints, potential inaccuracies in InsideAirbnb data and very few attributes in the dataset that can be related.

   Technical risks: Data integration, ETL pipeline failure, performance issues of data warehouse with big dataset.

   External validity: Results will not be likely to generalize across cities with diverse regulatory contexts.

6. **How much will this project cost?**

   In this project we are using NYC open data available to everyone, with open-source tools. We aim to keep the project efficient and at low-cost by using personal or university resources.

7. **How will you measure success along the way?**

   Currently, we aim to measure for two end points. For mid-presentation we aim to record data-preprocessing, error handling, ensure pipeline reliability by running basic queries to test the pipeline and structure the warehouse schema based on the complaints recorded in the NYC. In the final presentation, we aim to demo the project by presenting live queries and present visualizations on end-to-end analysis of 311 complaints in NYC.