CSCI 5622 Final Project Proposal:
Predicting Rental Market Changes from Airbnb Data for New York City

Group Members: Christina Cleveland, Maggie Jacoby

Motivation

Rental markets are complex systems with many unseen forces shaping them. Information asymmetry between owners and renters and the temporal structure of long-term leases means that property owners can be slow to respond to changes in the market. This makes rental trends difficult to predict or to witness in real-time.

We hypothesize that the decisions informing long-term rental markets have strong overlap with those informing short-term rental markets, such as Airbnb (see Figure 1). The ease of becoming an Airbnb "host" coupled with the rapid rate of occupant turnover means the Airbnb market reacts more quickly than the rental market to relevant contextual changes. Therefore, we propose that by observing the changes to short-term rental markets we can predict the changes that will occur in long-term rental markets.

Due to the complicated nature of the set of decisions influencing both the short (Airbnb) and long-term rental markets, changes to these markets are difficult to predict using heuristic models. However, this complexity makes the problem a good target for machine learning, in particular deep learning. Additionally, any interpretable model we could develop would be of great interest to real-estate investors, city-planners, and consumers alike.

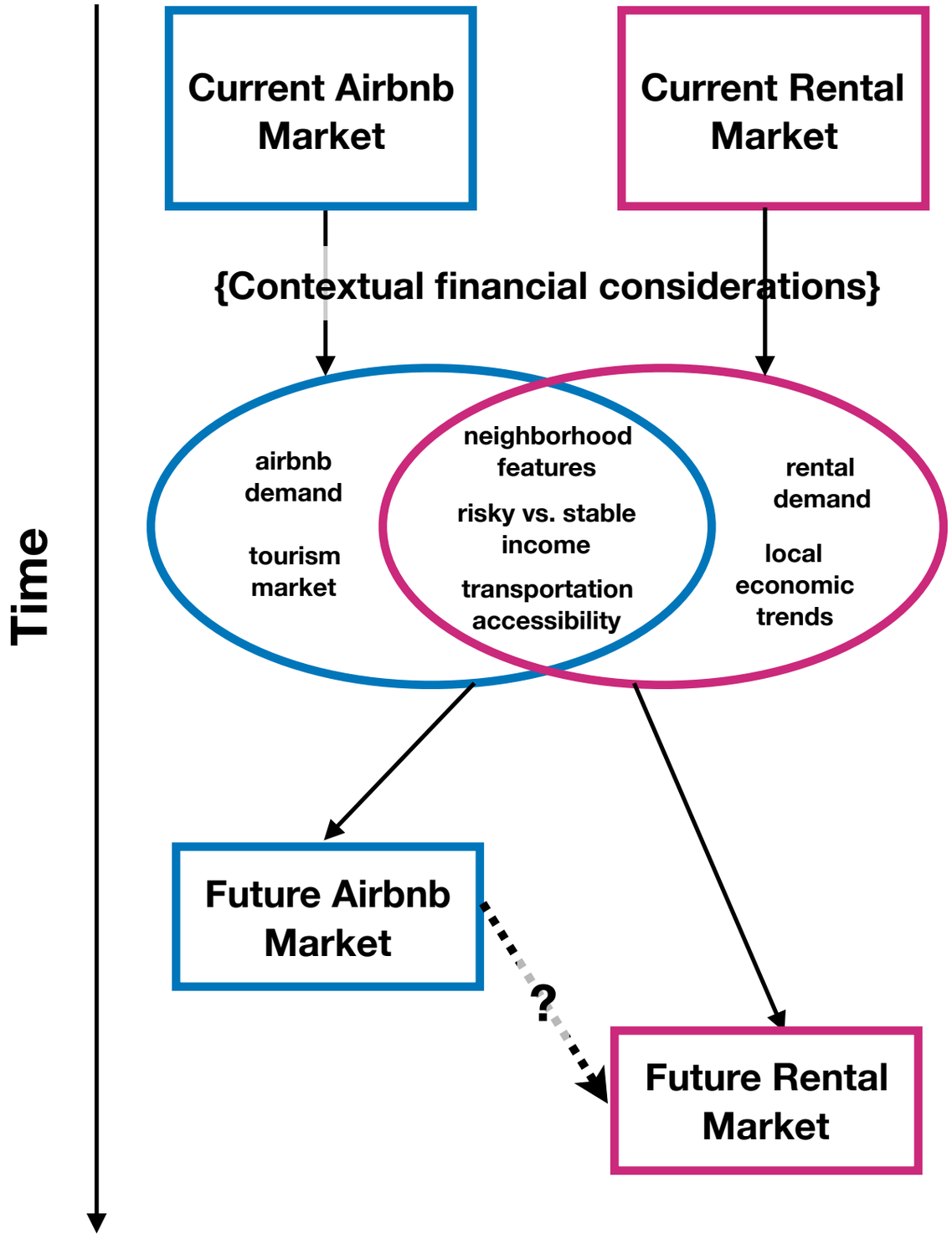

Figure 1: Hypothesized short-term and long-term rental market relationship

CSCI 5622 Final Project Proposal:
Predicting Rental Market Changes from Airbnb Data for New York City

Group Members: Christina Cleveland, Maggie Jacoby

Proposed Research Plan

For simplicity, we will study New York City alone, looking at it's most populous neighborhoods. We have access to the total set of Airbnb listings from New York City (the size of which is approximately 40,000 listings each with 96 features) for every month going back to January 2015. For the rental data, which will be used to label our Airbnb data, Zillow releases median rental prices by neighborhood and listing type per month going back to 2010.

We will start with exploratory analysis of the Airbnb data and a manual culling of features (see Figure 2). We will also use unsupervised techniques, *e.g.*, clustering techniques, for dimensionality reduction, feature selection and data labeling. The labels from the unsupervised methods will be compared to the labels assigned using the Zillow data set. Following the unsupervised methods, we will attempt to predict rent changes using a variety of machine learning methods, *e.g.*, random forests, deep learning, and Bayesian networks.

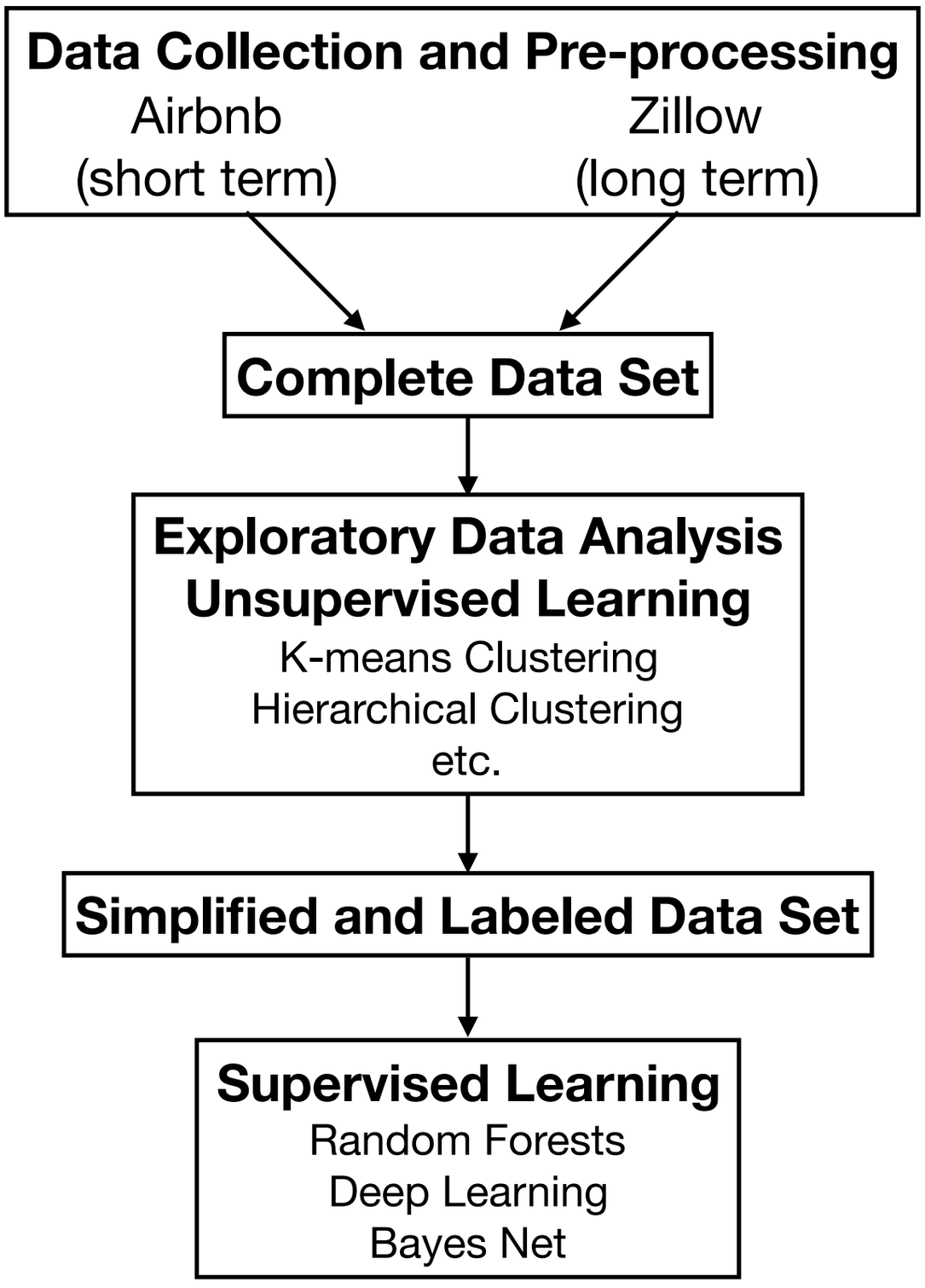

Figure 2: Proposed Research Pipeline