

**Course: DATA522 Solving Big Data Problems**

Project Topic

Date: July 13, 2021

Name: Wujian Xue

Student ID: 770031176

Content

[Project topic 3](#_Toc77089823)

[1.Team member 3](#_Toc77089824)

[2.Kaggle dataset 3](#_Toc77089825)

[3.Technologies and tools 3](#_Toc77089826)

[4.Project plan 3](#_Toc77089827)

[5.Reference 4](#_Toc77089828)

# Project topic

## 1.Team member

This project will be completed by Wujian Xue only.

## 2.Kaggle dataset

For this project, I will use FHFA House Price Indexes (HPIs) from Kaggle. It is located at <https://www.kaggle.com/tunguz/fhfa-house-price-indexes-hpis>

## 3.Technologies and tools

In this project, I plant to use following tools.

* R/RStudio
* Splunk
* Ambari (Hadoop File System)
* Github: <https://github.com/wx2123>

## 4.Project plan

House prices are not only important for businesses but also important for everyone. Just like stock indexes demonstrate the performance of stock market, the House Price Indexes (HPI) illustrate the changes in the housing market.

In this project, I will analyze Federal Housing Financing Agency (FHFA) House Price Indexes using tools such as R, Splunk, Ambari and Github. The HPI dataset has 10 columns and 114,731 observations. I will explore the dataset, group HPI by regions and visualize HPI.

Then, I will use machine learning models to predict the future house price indexes. I will start with some basic time series models such as AR, MA, ARMA. And then I will try other more complicated models.

Finally, I will compare results from different models and conclude the project. Further researches will also be dicussed.

## 5.Reference

Aderibigbe, T. & Chi, H. (2018). Investigation of Florida Housing Prices using Predictive Time Series Model, Proceedings of the Practice and Experience on Advanced Research Computing, July 2018, Article No. 92, Pages 1–4

Smola, A. & Vishwanathan, S.V.N. (2008). Introduction to Machine Learning, Cambridge Press.