**Sparkify churn project**

**Table of Contents**

1. Install
2. [Project Motivation](https://github.com/fxzero/Sparkify-Project#motivation)
3. [Files Description](https://github.com/fxzero/Sparkify-Project#files)
4. [Result](https://github.com/fxzero/Sparkify-Project#Result)
5. [Licensing, Authors, and Acknowledgements](https://github.com/fxzero/Sparkify-Project#licensing)

**Installation**

This project uses the following software and Python libraries:

Python:3.7.4

Pyspark:3.0.0

Pandas:0.25.1

Matplotlib:3.3.1

Seaborn:0.9.0

For this project, I worked in Jupyter workspace.

**Project Motivation**

This project aims to predict the user churn, then we can take some measure to reduce the churn rate, thus to get the benefits from this measure.

**Files Description**

We use the file (Sparkify20200801 .ipynb) to finish this project, this file contains the detailed process for load, data exploration, clean,build model and analysis.

**Result**

According to the results of the model, How long the user registered and whether they add more songs to playlist have the greatest impact. I post a blog about the detail, you can find it [here](https://lijianhua040704.github.io/).

**Licensing, Authors, Acknowledgements**

Must give credit to Udacity for the project. You can't use this for you Udacity capstone project. Otherwise, feel free to use the code here as you would like!