**Readme.md**

**Sparkify project**

**Table of Contents**

1. [Installation](https://github.com/fxzero/Sparkify-Project#installation)
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 3.6

Spark

Pyspark 3.0

Pandas 0.19

Matplotlib

Seaborn

You will also need to have software installed to run and execute a Jupyter Notebook.

If you do not have Python installed yet, it is highly recommended that you install the Anaconda distribution of Python, which already has the above packages and more included. And for Spark, you can do this using AWS or IBM Cloud.

**Project Motivation**

This is udacity's capstone project, using spark to analyze user behavior data from music app Sparkify.

Sparkify is a music app, this dataset contains two months of sparkify user behavior log. The log contains some basic information about the user as well as information about a single action. A user can contain many entries. In the data, a part of the user is churned, through the cancellation of the account behavior can be distinguished.

**Files Description**

**Sprakify .ipynb** Main file of the project, it demonstrates the process of using pyspark to explore the data and build the model.

**Result**

According to the results of the model, it is the frequency of Thumbs Down that has the greatest impact. Churn users have more Thumbs Down. Naturally, users will leave if they are not satisfied.

I post a blog about the detail, you can find it [here](https://medium.com/@fxzero/how-to-predict-user-churn-using-pyspark-fe25f6de1d7a).

**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!