

Project Name: Fitbit Time Series Project

Project Type: optional solo project

Project Goals:

Use explore time series data and use time series algorithms to predict next 2 weeks of fitbit data

Construct a ML time series model that predicts next 2 weeks of data with accuracy better than baseline

Deliverables: Main Notebook of Pipeline process with summary and conclusions, csv of next 2 weeks data

Stage	Tools	Brief Description of Process	Challenge Resolution
Plan	<ul style="list-style-type: none">Visual Studio	<ul style="list-style-type: none">In Visual Studio create a new readme.md file to outline project plan, create GitHub repoImport key elements and deliverables from curriculum requirements	<ul style="list-style-type: none">No unusual challenges in this section
Acquire	<ul style="list-style-type: none">Visual StudioGoogle Sheets.py script	<ul style="list-style-type: none">Data received in csv files included multiple tables in one sheet and separate files for each weekUsed Google Sheets to initially investigate and rearrange data into one csv file for analysis	<ul style="list-style-type: none">Once I realized the set up of the csv files were the issue resolving that using spreadsheet program was not an issueDecided not to use a programmatic approach for this one off issue
Prepare	<ul style="list-style-type: none">Google SheetsJupyter Notebook	<ul style="list-style-type: none">Prepared data formats in spreadsheet as well as column names and titlesCondensed data into one table for import into pandas	<ul style="list-style-type: none">No unusual challenges in this section
Explore	<ul style="list-style-type: none">Jupyter NotebookSeaborn	<ul style="list-style-type: none">Used pandas time series resampling and plotting features to investigate dataDid not find a seasonal/cyclic trend in the	<ul style="list-style-type: none">Used examples from lessons

	<ul style="list-style-type: none"> Matplotlib 	<ul style="list-style-type: none"> data Did find a slight upward linear trend 	
Model	<ul style="list-style-type: none"> Jupyter Notebook Sklearn Multiple ML models tested 	<ul style="list-style-type: none"> Used last observed, and simple average models to determine a baseline (selected simple average as baseline to beat). Used rolling average and Holts models for predictions Unable to use previous cycle because data did not contain a cycle 	<ul style="list-style-type: none"> Used examples from lessons
Evaluate	<ul style="list-style-type: none"> Jupyter Notebook Sklearn 	<ul style="list-style-type: none"> Determined rolling average was most accurate based on least RMSE Using 7 day rolling average produced model that was 28% better than simple average baseline 	<ul style="list-style-type: none"> Used examples from lessons
Model Explanation	How does your algorithm work?	<ul style="list-style-type: none"> The rolling average takes the average over the period specified moving the window one increment at a time 	<ul style="list-style-type: none"> No unusual challenges in this section
Delivery	<ul style="list-style-type: none"> Jupyter Notebook 	<ul style="list-style-type: none"> Added anchor link to jump to conclusions at bottom, also utilized Table of Contents to organize and jump to sections 	<ul style="list-style-type: none"> No unusual challenges in this section