README.md



Introduction to Pandas and Matplotlib

pandas is a powerful, open source Python library for data analysis, manipulation, and visualization. matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. The goal of the following Jupyter notebooks is to give you some concrete examples for getting started with pandas and matplotlib.

Jupyter Notebooks

- DataFrames and Series
- Sorting, Filtering and Applying
- Data Cleaning and Preparation
- Data Aggregation
- Data Visualization

Practice Problems

- Problem 1 Solution
- Problem 2 Solution

Homework (due Sept 11)

- Stanford Open Police Dataset
- US Baby Names

Datasets

Filename	Description	Source
yob1880.csv yob2018.csv	frequency of baby names from 1880 through 2018	SSA

Filename	Description	Source
police_ri.csv	Data on vehicle and pedestrian stops from law enforcement departments across Rhode Island	Stanford Open Policing Project
sales.csv	Monthly Sales Data of sell of electronics equipment across USA in 2019	Kaggle
Fremont.csv	Fremont bridge bicycle counter	City of Seattle Open Data portal
drinks.csv	Alcohol consumption by country	FiveThirtyEight
imdb_1000.csv	Top rated movies from IMDb	IMDb
titanic.csv	Famous Titanic daseset from Kaggle	Kaggle
ufo.csv	Reports of UFO sightings from 1930-2000 (the truth is out there)	National UFO Reporting Center
chipotleorders.csv	Online orders from the Chipotle restaurant chain (link to The Upshot article)	TheUpshot
advertising.csv	Sales info in 200 different markets, along with advertising budgets for three different media: TV, radio, and newspaper	An Introduction to Statistical Learning
tips.csv	Restaurant tipping	Python for Data Analysis