

Time Series Data Visualization Project Overview

Project Summary

This project involves creating comprehensive visualizations of time series data to analyze patterns in freeCodeCamp.org forum page views from May 2016 to December 2019. The project demonstrates proficiency in data cleaning, manipulation, and visualization using Python's core data science libraries.

Objectives

- Analyze and visualize time series patterns in web traffic data
- Identify yearly and monthly growth trends
- Create multiple visualization types to reveal different aspects of the data
- Implement data cleaning techniques to improve data quality

Dataset Information

- **Source:** fcc-forum-pageviews.csv
- **Time Period:** May 9, 2016 to December 3, 2019
- **Data Points:** Daily page view counts for freeCodeCamp.org forum
- **Key Field:** Date-indexed page view counts

Technical Stack

- **Python Libraries Used:**
 - **Pandas:** Data import, manipulation, and cleaning
 - **Matplotlib:** Line chart and bar chart creation
 - **Seaborn:** Statistical visualization and box plots

Data Processing Pipeline

1. Data Import and Preparation

- Import CSV data using Pandas
- Set date column as index for time series analysis
- Ensure proper datetime formatting for temporal operations

2. Data Cleaning

- Apply statistical filtering to remove outliers

- Filter out extreme values (top 2.5% and bottom 2.5% of dataset)
- Maintain data integrity while removing anomalous readings

3. Data Analysis and Visualization

Visualization Components

Line Chart (`draw_line_plot`)

- **Purpose:** Show daily page view trends over time
- **Title:** "Daily freeCodeCamp Forum Page Views 5/2016-12/2019"
- **Axes:** Date (x-axis) vs Page Views (y-axis)
- **Insights:** Reveals overall growth trends, seasonal patterns, and daily fluctuations

Bar Chart (`draw_bar_plot`)

- **Purpose:** Compare average monthly page views across different years
- **Structure:** Grouped bar chart with years on x-axis
- **Legend:** Monthly breakdown with "Months" title
- **Axes:** Years (x-axis) vs Average Page Views (y-axis)
- **Insights:** Highlights year-over-year growth and seasonal monthly patterns

Box Plots (`draw_box_plot`)

- **Purpose:** Statistical distribution analysis
- **Structure:** Two adjacent box plots
 1. **Year-wise Box Plot (Trend):** Shows annual distribution patterns
 2. **Month-wise Box Plot (Seasonality):** Reveals seasonal variations
- **Insights:** Identifies data spread, outliers, and comparative distributions

Key Skills Demonstrated

Data Science Fundamentals

- Time series data handling and indexing
- Statistical outlier detection and removal
- Data aggregation and grouping operations

Visualization Techniques

- Multi-plot figure creation
- Custom styling and labeling
- Statistical plot interpretation
- Color schemes and legend management

Python Programming

- Function-based code organization
- Library integration (Pandas, Matplotlib, Seaborn)
- Data manipulation workflows
- File I/O operations

Project Outcomes

Analytical Insights

- **Growth Trends:** Quantified forum engagement growth over 3+ years
- **Seasonal Patterns:** Identified monthly usage variations
- **Data Quality:** Improved dataset reliability through outlier removal

Technical Deliverables

- Clean, reusable visualization functions
- Comprehensive data cleaning pipeline
- Multiple chart types for different analytical perspectives
- Professional-quality data visualizations

Business Value

- **Web Analytics:** Demonstrates ability to analyze web traffic patterns
- **Growth Analysis:** Shows capability to identify and visualize business growth
- **Data Quality:** Proves skills in data cleaning and preparation
- **Communication:** Creates clear, informative visualizations for stakeholders

Future Enhancements

- Interactive dashboard development
- Predictive modeling for future page views
- Correlation analysis with external factors

- Real-time data pipeline integration

Technical Specifications

- **Input Format:** CSV with date and page view columns
- **Output:** Three distinct visualization functions
- **Data Range:** 3.5+ years of daily observations
- **Filtering:** 95% data retention after outlier removal

This project showcases end-to-end data analysis capabilities, from raw data processing to publication-ready visualizations, making it an excellent demonstration of practical data science skills in web analytics and time series analysis.