

Data Visualization

Introduction to Data Analysis with Python - Spring Semester 2026

Cecilia Graiff

March 20, 2026

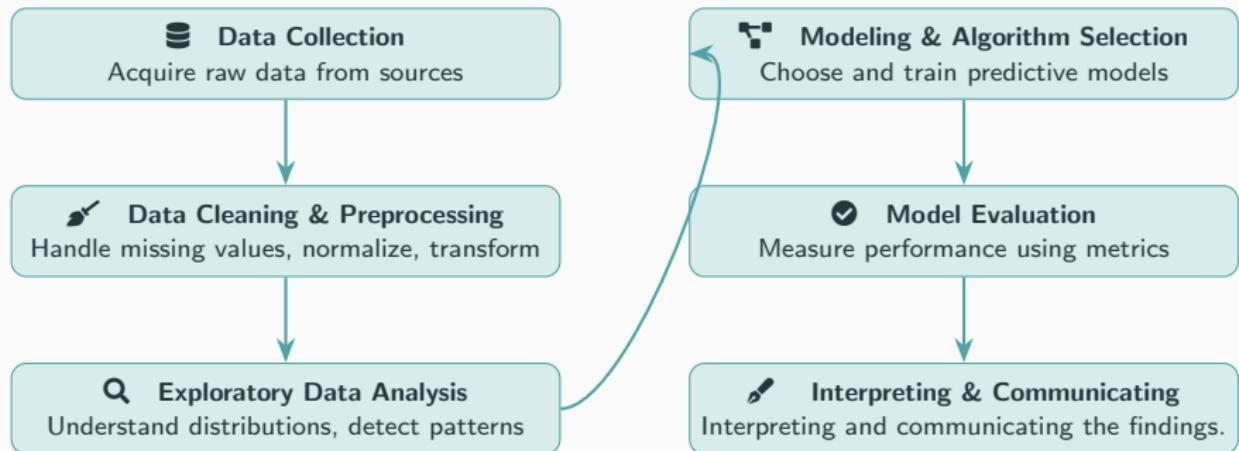
ALMAnaCH, Inria Paris - École d'Affaires Publiques, SciencesPo
cecilia.graiff@sciencespo.fr

Table of Contents

1. Homework Correction
2. Simple Data Visualization
3. Advanced Visualization Techniques

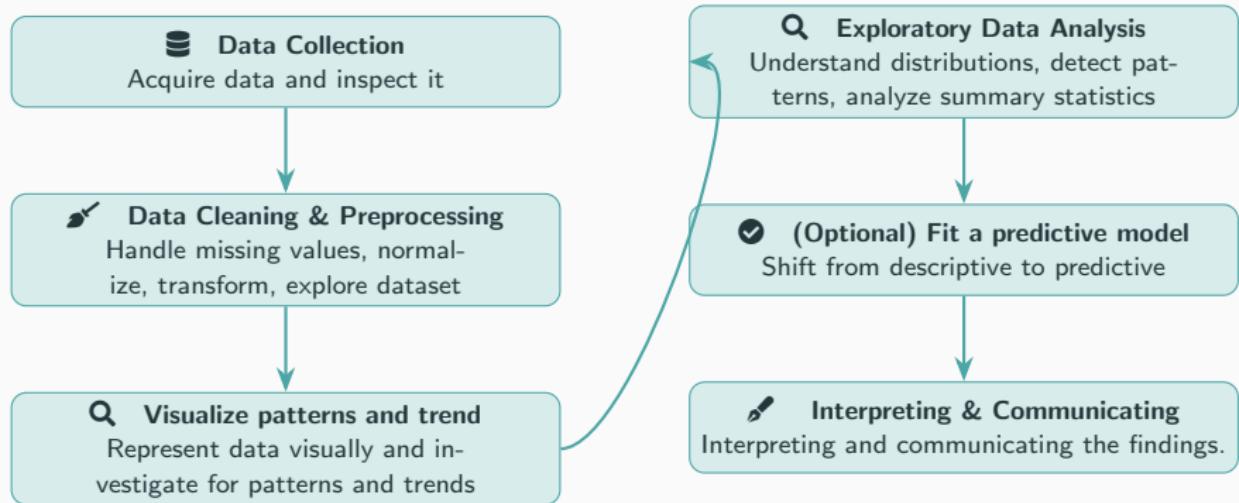
Homework Correction

Recap: Data Analysis Pipeline



- **Advanced data analysis** shifts from descriptive to predictive
- This introductory class is mostly focused on steps 1-3

Possible pipeline for the group project



Course Overview

- Up to now: gaining necessary Python skills
- Next lessons: Basic statistics for data analytics, including some theory

Simple Data Visualization

Scatterplot

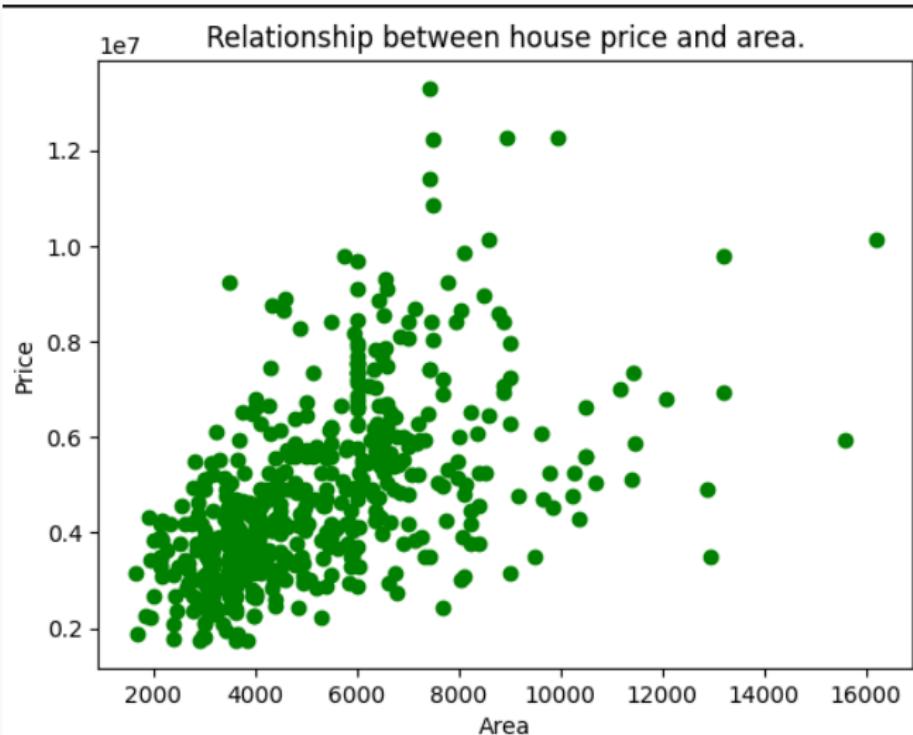


Figure 1: Scatterplot representing the relation between house prices and area.

- Displays the relationship between two numerical variables using dots.
- Each point represents an observation's values on the x- and y-axes.
- Useful for identifying correlations, patterns, or outliers.

Bar Plot

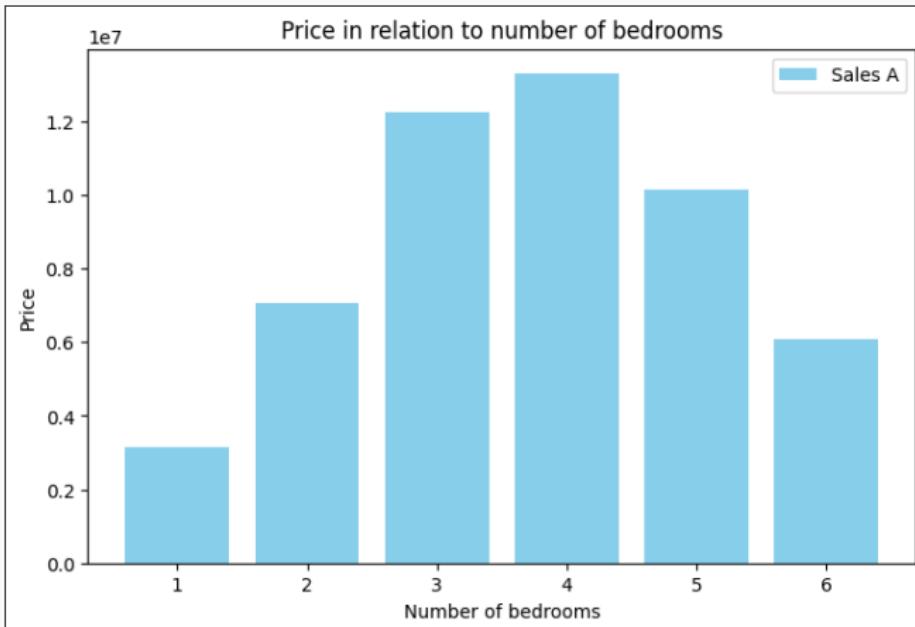


Figure 2: Bar plot representing price value per number of bedrooms.

- Displays categories on one axis, values on the other
- Can be horizontal or vertical
- Useful for comparing **discrete values**

Histogram

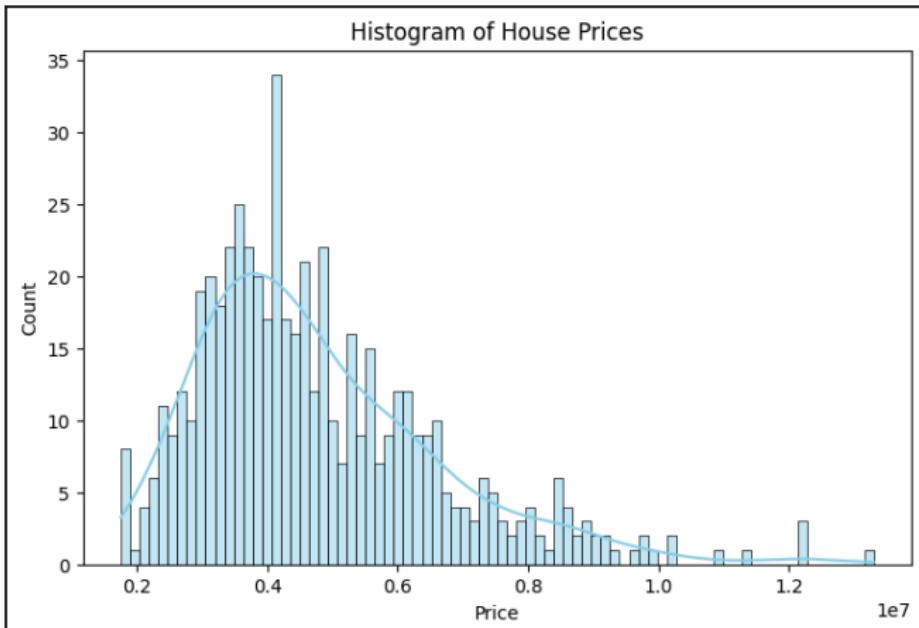


Figure 3: Histogram representing the continuous variable price.

Bar Plot vs Histogram

- Histogram represents a **continuous value**, bar plot represents **frequency amounts for each value**
- Histogram: Divide value into bins, and plot those bins
- Bar plot: plot real number of each value
- Purpose of bar plot: frequency count
- Purpose of histogram: checking data evolution

Advanced Visualization Techniques

Summary

- Heatmaps
- Pairplots
- Boxplots
- Violin Plots
- Facet Grids
- Grouped Bar Plots
- Regression
- Time-Series (optionally: with Rolling Averages)

Heatmap

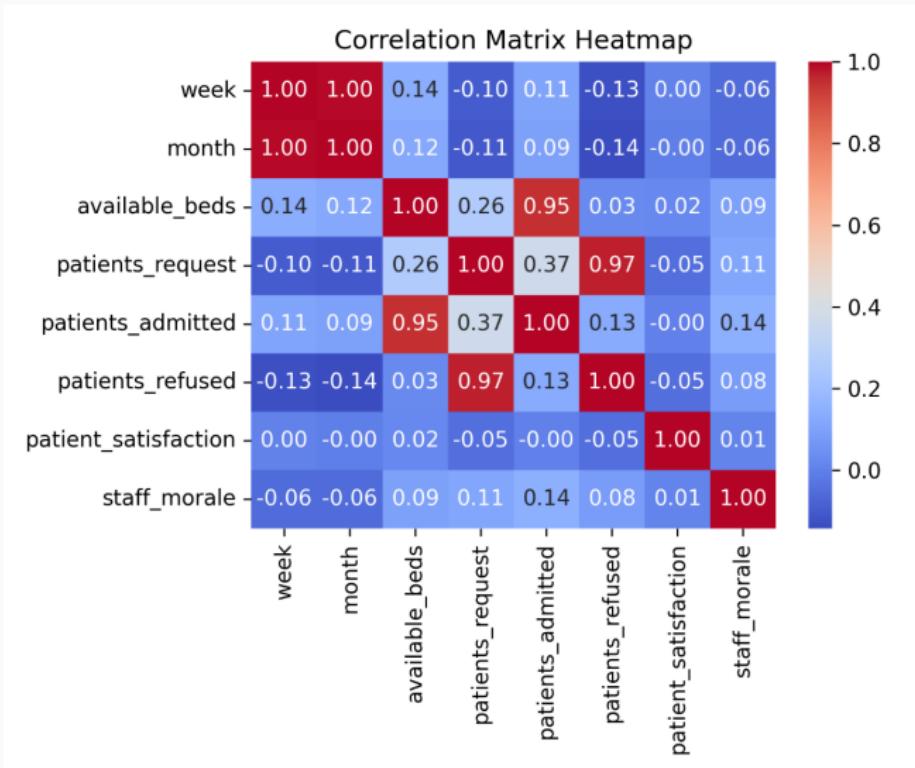


Figure 4: Correlation matrix heatmap of numerical features

- Plots the values one against the other in a matrix format
- The number you see is the **pairwise correlation coefficient**
- Quantifies linear relationship between the variables
 - 1: perfect linear correlation
 - 0: no linear correlation
 - -1: perfect negative linear correlation
- Useful to spot correlations among the variables
- More info about correlation to come in the EDA lecture

Pairplot / Scatter Matrix

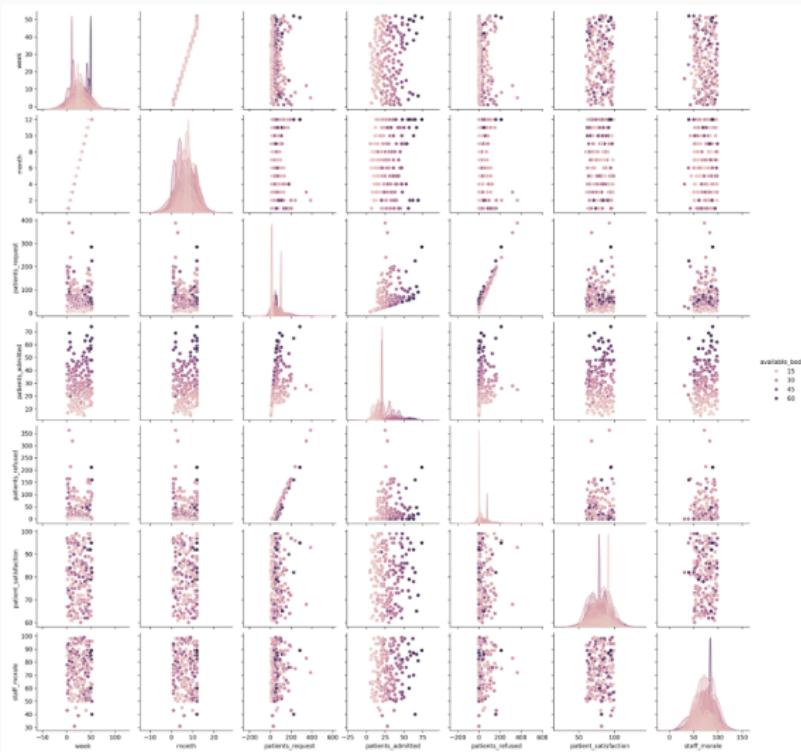


Figure 5: Pairwise relationships between variables

- Similarly to heatmap, plots correlation between variables
- In form of a plot and not a number

Boxplots

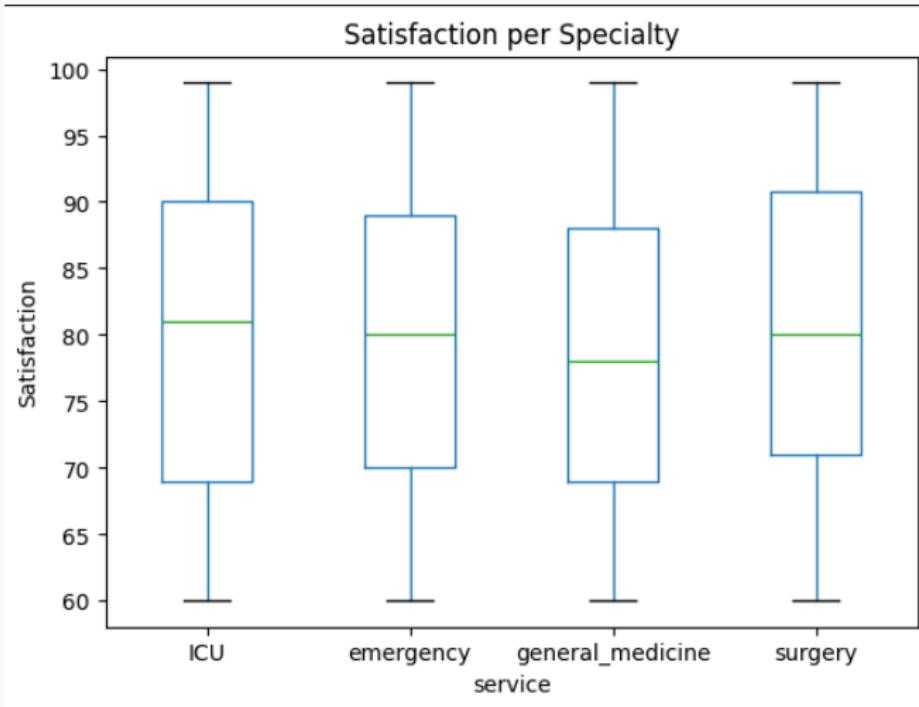


Figure 6: Boxplot representing patient satisfaction per category.

- Box: Interquartile range (values between 25% and 75%)
- Line: Median
- Upper and lower whisker, eventually outliers

Violin Plot



Figure 7: Distribution of patient satisfaction

- Wide sections = many points (high density)
- Narrow sections = few points (low density)
- Inside: data points, eventually median and IQ-range

Facet Grid / Small Multiples

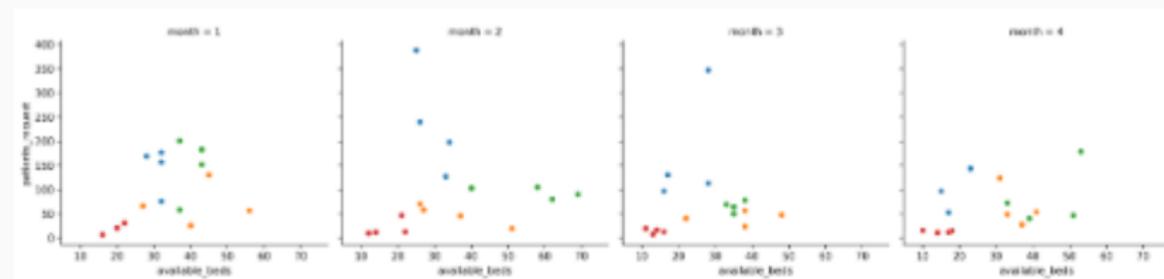


Figure 8: Scatter plots of available beds vs patients request per service.

- Same kind of plot (here: scatterplot) across multiple subsets
- Useful to compare patterns across subsets
- Parameters: rows (x-axis), column (y-axis), hue (colour, optional)

Stacked / Grouped Bar Plot

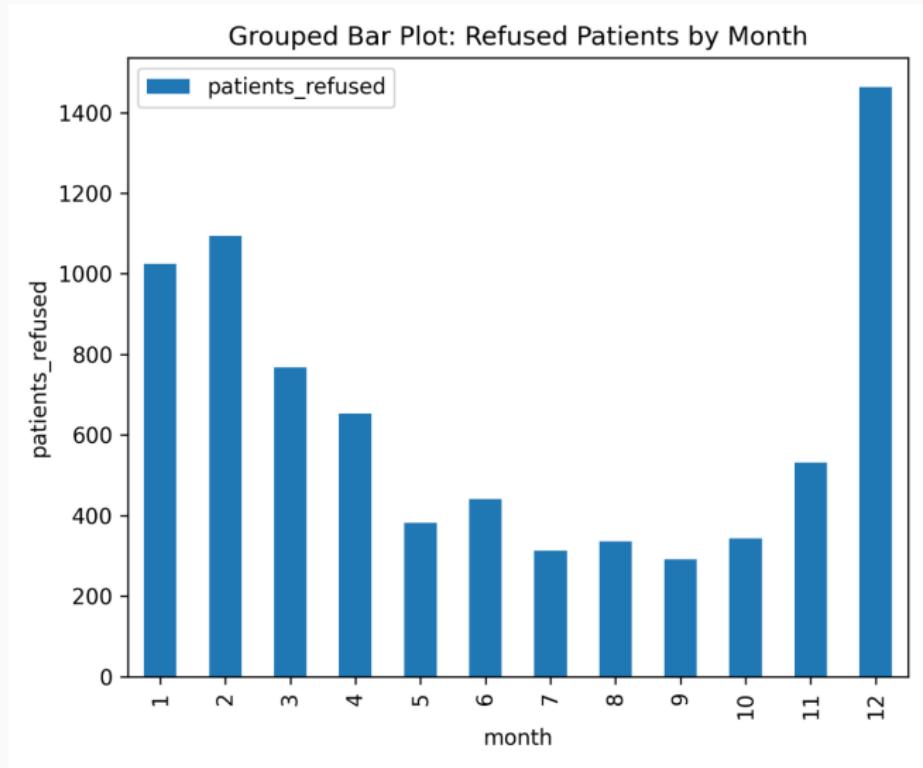


Figure 9: Enter Caption

- Same idea as bar plot, but more advanced: grouping several variables
 - Here: refused patients by month **for all categories**

Regression / Trend Lines

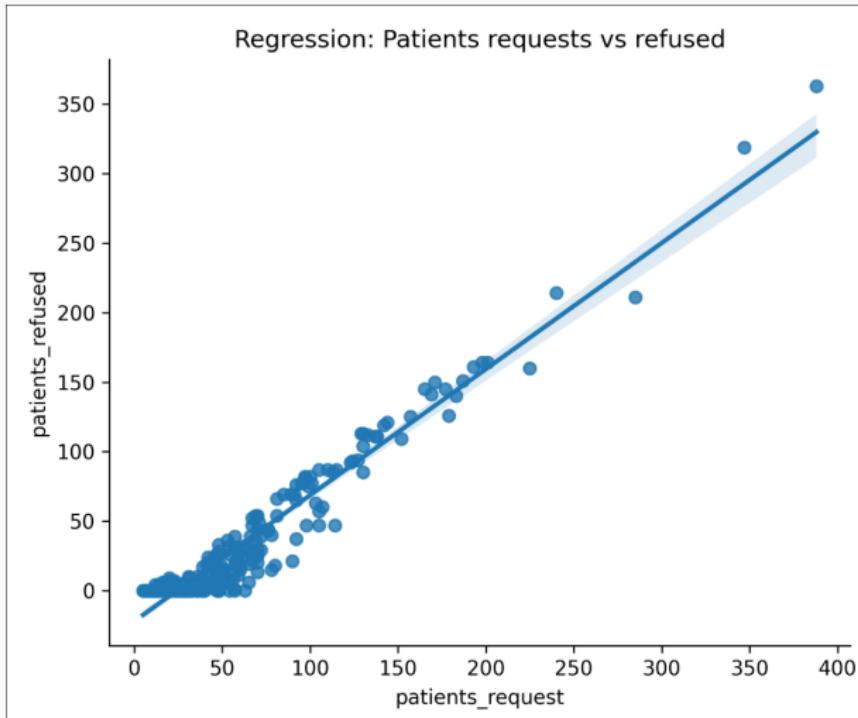


Figure 10: Regression lines on a scatter plot.

- Plots linear relation between two variables
- Useful to detect trends
- Eventually, add hue parameter

Time-Series with Rolling Averages

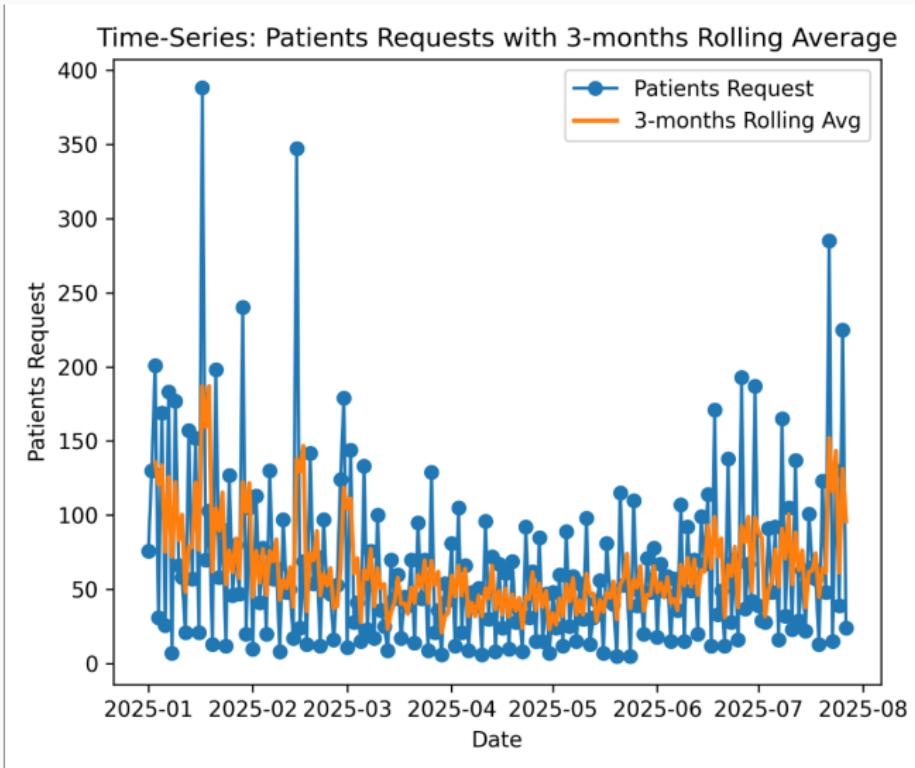


Figure 11: Monthly patients admissions with 3-months rolling average

- Data in relation to time (here: months)
- Data taken at **equal points in time**
- Useful to analyze temporal structure and temporal changes of data
- Rolling average: smooth monthly data (here: three months, so average of each month with 2 previous months) to highlight trends without being influenced by possibly unimportant changes

Summary

- Heatmaps → correlations and patterns
- Pairplots → multi-variable relationships
- Violin plots → distributions with density
- Facet grids → subgroup patterns
- Stacked/Grouped bars → categorical comparisons
- Regression → trends and group differences
- Time-series → temporal trends and smoothing