Analyzing and Visualizing Ridership Patterns in Île-de-France Rail Network

Author

Etienne Côme

Published

November 12, 2024

Project Overview

In this data analysis project, students will delve into the ridership data of Île-de-France's railway stations spanning the years 2018 to 2023. The primary objective is to analyze and visualize the ridership patterns, creating a dashboard that allows stakeholders to monitor and compare ridership against the norm. The analysis will specifically focus on discerning variations from a typical week, distinguishing between regular weeks and holiday periods.

Project Tasks

1. Data Collection and Cleaning

- The data that you will use for this study are available trought the STIF open data portal :
  - [validation per day and stop point first semester of 2023](#)
  - [validation per day and stop point 2015 - 2022](#)
  - [stops i.e "Zone d'arrêt" locations](#), for this last file the spatial data file can be downloaded directly at https://eu.ftp.opendatasoft.com/stif/Reflex/REF_ZdA.zip
  - other ressources from stif open data website
- Gather and compile ridership data from Île-de-France railway stations for the specified period, starting from from 2018-01-01.
- Clean the dataset to handle missing values, outliers, and any inconsistencies.
- Aggregate the dataset at the "Zone d'arrêt" level defined by the ID_REFA_LDA feature, collect geographical data about their localisations
- Provide a reproducible R script cleaning.R that gather all the cleaning steps

2. Exploratory Data Analysis (EDA)

- Conduct exploratory analysis to identify overall trends and patterns in ridership.
- Explore seasonality, monthly trends, and potential outliers affecting the data.

4. Comparison with Norms

- Define a baseline "normal" week and investigate deviations during holiday and non-holiday periods.
- Assess the impact of vacations and school breaks on ridership patterns.

5. Dashboard Development using Shiny

- Build an interactive dashboard using the Shiny framework in R.
- Include key visualizations depicting overall ridership trends, weekly variations, and comparisons with the norm.

Example of tools :

- Allow the user to select a reference period and a period to compare against and provide meaningfull figures to highlight the differences between the two selected periods, discriminating by day of week.
- Allow the user to quickly select a stations of interest with a webnmap and procides key statistics about the current trend for this stations.

6. Statistical Methods

- Apply appropriate simple statistical methods to validate findings and draw meaningful insights from the data.
- Use statistical tests to assess the significance of observed variations, when possible.

7. Report and Documentation

- Produce a comprehensive report detailing the entire analysis process, methodologies used, and key findings.
- Include insights into how the railway network's ridership has evolved over the studied period.

8. Shiny App Deployment

- Deploy the Shiny dashboard on shinyapp.io for accessibility and user interaction.
- Ensure the application is user-friendly and capable of providing dynamic insights into ridership patterns.

Deliverables

- A commented R script that collect all the cleaning / preprocessing steps that were developed
- A well-documented report in quarto format including code, analysis, and interpretations.
- A functional Shiny dashboard accessible through shinyapp.io.

Dead line

8 January 2024 by e-mail

Group Size

3 to 4 students per group.

Note: The project encourages collaborative teamwork, critical thinking, and the application of data analysis techniques to real-world transportation data.