CityBike Lisbon: Location Data Modelling with Python

Final project – Statistical Modelling with Python

This project explores how bike availability at Lisbon's CityBike stations relates to nearby businesses, such as bars, cafés, and restaurants, as well as their average rating and popularity.

Primary Goal
Identify which location characteristics best explain variation in free bikes and total docking slots at Lisbon CityBike stations.

Data Timestamp:
12:39 AM Sunday, July 27 2025

Key Insights

Independent variables

- Category type (restaurant, bar, café, coffee)
- Average rating
- Average popularity

Dependent variables

- Free bikes
- Total bike slots

Location density & API bias

Foursquare returns at most 50 venues per station, which biases dense areas. To reduce this, only venues within **298 m** of each station were retained

Key findings

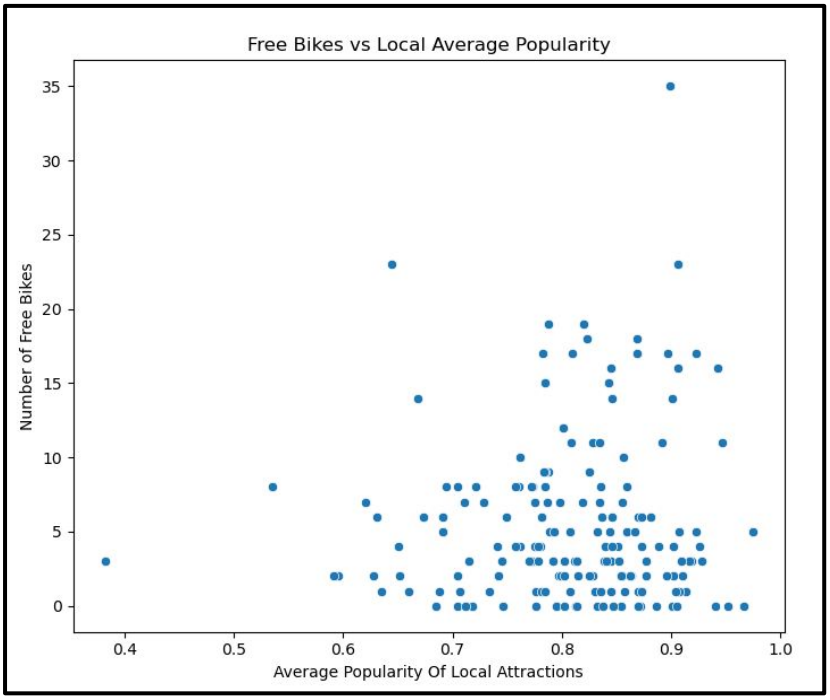
Location category type (e.g., restaurant, bar, café) is a **weak predictor** of free bikes or total slots.

Average rating and popularity are also **poor predictors** of station metrics

Overall, no strong predictive relationships were observed

However, the counts of **bars** and **cafés** near a station provide the strongest available signal, jointly explaining **≈18%** of the variance in free bikes

Details



Free bikes vs average local popularity for each station

Next Steps

Additional variables to collect

- Time of day
- Day of week / season
- Longer-term patterns (daily, weekly, monthly trends)

Further feature engineering

- Group venues into broader categories (e.g., public/government, retail, hospitality)
- Test whether these broader groups improve model performance