To-Do

Alberto Calabrese

2024-05-13

EDA

This part is dedicated to the explorative data analysis of the data loaded above

The aim is to understand the data structure and the main characteristics of the data

1. Understand the data structure
2. Understand the main characteristics of the data
3. Compare the data between the different categories
   - Cinema & Teatro
4. Understand the trend of the data by month
   - Cinema
   - Teatro
   - Generale
5. Understand the correlation between the different areas
6. Understand the correlation between the different activities
7. Understand the correlation between the places and the activities
8. Understand the correlation between the places and the entries (spettatori)
9. Understand the correlation between the expenses and entries
10. Understand the correlation between the number shows and ????
11. Understand the correlation between the different macro areas

MODELS

Regression and model part between the regional data (Cinema, Teatro, Generale, Organizzatori, Luoghi)

1. Creazione modelli
   - Regressione lineare
   - Regressione lineare multipla
   - Regressione logistica
2. Selezione modello
   - Forward
   - Backward
   - Stepwise
   - BIC - AIC - ecc. . . .
3. Analisi residui
   - Test
   - Leverage
   - Collinearità
   - Analisi degli outliers
4. Forecasting . . . .

Idea:

Classificazione di un cinema per definire se è efficiente ovvero il rapporto spesa spettatori è maggiore della media