IONOS

Data Engineer - Data Challenge

Titanic Survival prediction

Dataset

Download the titanic dataset from kaggle https://www.kaggle.com/competitions/titanic/data or github (e.g. https://github.com/datasciencedojo/datasets/blob/master/titanic.csv)
Please reach out to Benjamin Scheer (benjamin.scheer@ionos.com) if you encounter any issues getting your hands on the correct dataset.

The challenge

1. Develop an ETL pipeline including at least these steps:
    - load the data from the csv
    - perform the necessary preprocessing / cleaning of the data
    - Feature engineering (e.g. extracting title from names, normalization of numerical values)
    - save the transformed data into a database (e.g. SQlite)
2. Write a script (preferably in Python) to automate that ETL process and propose a way how to run this script daily (assuming there is an updated titanic-dataset every day)
3. Design a simple data pipeline architecture diagram (can be hand-drawn or created using diagram software) that includes:
    - Data ingress (e.g., raw CSV files)
    - Data processing (ETL steps)
    - Data storage (e.g., SQL database)
    - Machine learning model training and validation
4. Provide a compact analysis of the provided data
5. provide the SQL to answer the following questions from your database:
    - What's the average age of women that survived the sinking of the titanic?
    - What are the average and maximum fares for each class?

Submission

Provide your work in a git repository or comparable format. If you use github, please provide access to the repository to Benjamin Scheer (invite benjamin.scheer@ionos.com as a collaborator) and send a mail including your name & github account name to him, so we can match the identities.
⇒ Include a brief documentation of your approach, implementation details
⇒ If your github account name does not include your real name, please leave us a hint so we can assign your submission correctly.
The target group & reviewers are your interviewers from IONOS which are Data Scientists, Data Engineers or Software Engineers.