Data Hackathon – a.y. 2021/2022

Prof. A. Longo, Prof. A. Martella, Prof. M. Zappatore

Data for Ecological Transition – Using open data to assess EU renewable energy context

UN Agenda 2030, Italian PNRR and the increase of our energy bills demand for a general awareness towards ecological transition to create a more sustainable world.

Open data is one of the resources available to understand our reality, analyse the context and support decision making.

THE MISSION

The aim of the contest is to provide well-grounded evidence and support decision makers on how to improve the exploitation of renewable energy in the EU . Therefore, it is requested to

1. Analyze a big dataset describing renewable power plants in different EU countries by reverse engineering and querying it.
2. Transform the dataset in order to support situation assessment and decision making by developing adequate analytics and data dashboards.

TASKS

Task #1: Searching and analysing the dataset

You are provided with a multi-file dataset describing renewable power plants in different EU countries. Each country provides a .csv file whose structure may slightly differ from or overlap to the structure of the same file provided by another EU country. Therefore, you must analyse and reverse engineer the dataset at first according to the list of steps suggested below:

- Download the dataset from

  https://data.open-power-system-data.org/renewable_power_plants/2020-08-25

- Describe the subject of the dataset, by providing motivations to support your statements
- Reverse engineer the dataset, by designing its conceptual model. The conceptual model (i.e. an ER or EER diagram) must cover the whole datasets. Comment your modelling choices and highlight peculiarities (positive and negative aspects encountered when reverse engineering the dataset). Data dictionary is fundamental to clearly describe your model. Analyse the coverage of the intensive and extensive model against the given datasets
- Analyse the quality of the dataset: for each information object, both in intension and extension, report any potential issue in data values, formats (e.g., NULL values, differences in value

formats – i.e., dates, addresses, etc.), and errors in values and/or not respected model constraints.

- Implement the following minimum number of queries:
    a. Analyse each relation itself (i.e. for each relation report the number of tuple) and in relationship with the others (i.e. for each referential integrity constraint count the number of tuples),
    b. For each relation list the attributes with the corresponding number of null values, eventually ranked from the maximum to the minimum
    c. For each relation list the range of values (or the number of different values) of each attribute domain

Task #2:

Suppose you are a Chief Energy Officer from the European Commission who must support the EU Renewable Energy Commission to define the Renewable Energy strategic plan! It means providing evidence about how renewable power plants are managed and distributed in the EU

You want to assess specific aspects, such as energy source types by region, electrical capacity in MW by country, power plant types by technology, usage of wind turbines, average plant age by country, etc.

Therefore, the following list of steps is suggested:

- Develop the goal/stakeholder models and the related use case diagrams for navigating into data (at least 3)
- Design the DFMs and the related star schemas/snowflake schemas (at least 3)
- Using a dedicated BI platform or a data visualisation solution (e.g., Microsoft Power BI, Apache Knime, Flourish, etc.), develop the dashboards related to the use cases according to the DFMs

RULES:

- Each group is made up of 4 students at maximum
- The ranking goes from 1 (worst, just achievable by taking part to the contest) to 6 (best, assigned to the best group(s) only). The scores will be considered by each class lectures at the final evaluation (i.e. for big data management, data management and big data and CRM for business decision making students will gain the scores on the final evaluation, etc.).
- Each group will be evaluated according to
    a. Inter-discipline of team
    b. Completeness of the work
    c. Level of details achieved in the analysis

d. Additional data and queries (number and complexities) implemented to extend and explore the dataset and their effectiveness to browse into data

e. Number of DFMs modeled and developed (at least 3)

f. Number of use cases designed and developed

g. Communication skills (i.e., quality of presentation and capacity of supporting statements)

TIMELINE:

· Group creation: each student interested in attending the contest must submit the form at the following link by Dec. 15th.

https://docs.google.com/forms/d/e/1FAIpQLSf242xePy7ZaBwDFenEVLailq-RUx6K85l3el1LyNr-8fCBtw/viewform

On the 20th Dec. at the kick off meeting will be held online at a provided Teams link.

· Deadline for the contest: 5 weeks from January 3rd to February 7th, 2022

· Intermediate meetings for clarifications: January 10th and January 24th

· Final presentation: Feb 14th. 2022

CONSTRAINTS

· Use MySQL DBMS and auxiliary libraries for analysing the dataset

· In the report provide the structure of the queries and the results

· The report can be a presentation or a document

· Release the .sql script of the queries in the report

· In the report provide the DFM, the use cases and the BI tools project files.