**BTP GROUP – DOMAIN SELECTION**

*Project for the course Graph Databases 2024-25*

**GROUP MEMBERS:**

|  |  |  |  |
| --- | --- | --- | --- |
| NAME | SURNAME | EMAIL | ID NUMBER |
| Ali | Bakhshayesh | [ali.bakhshayesh@studenti.unipd.it](mailto:ali.bakhshayesh@studenti.unipd.it) | 2071502 |
| Francesco | Chemello | [francesco.chemello.1@studenti.unipd.it](mailto:francesco.chemello.1@studenti.unipd.it) | 2121346 |
| Marco | Martini | marco.martini.7@studenti.unipd.it | 2087642 |

**TOPIC:**

This Project focuses on the traffic flow using the datasets from [Bologna's Open Data](https://opendata.comune.bologna.it/pages/home/) platform.

This dataset has been chosen to represent a real-world scenario through a graph database.

This is particularly interesting since, while data analysis is available and has already been done in the open data platform, putting everything together in a simple-to-query database to infer more data through connected graphs is something we deemed useful for the community.

We deemed it especially important because we wanted to contribute to the *"Città 30"* project linked [here](https://www.bolognacitta30.it/) and see whether the choice of the infrastructure changes aligns with the data we have for future work we will be able to import new data and compare the results to previous years easily.

**CHALLENGE AND CHARACTERISTICS:**

The main challenge is linking everything together. For that, we plan to use the *"Archi stradali"* datasets that include all the roads and street infrastructures, which are linked through nodes, creating a graph on itself.

Despite this, linking everything is particularly tricky, since the data is not always directly correlated, but we're confident that due to the nature of a graph database, we can easily connect different topics through other common grounds.

An example could be the pollution sensors, which don't have a clear street location because of the way they encode it, but by implementing a good parser we can link it to existing roads.

We also had to consider which years to model as we don't always have the latest data for some tables (for example accidents data), and it can get big for what concerns storage size very fast, which would be outside the scope of the project. It also must be somewhat relevant to the *"Città 30"* project so the data can't be too old either, for that reason we've chosen datasets from 2019 to 2022.

The largest datasets are the traffic control *"Rilevazione flusso veicoli tramite spire"* and pollution *"Dati centraline qualità dell'aria"* as they provide daily data about it being around 300MB and 100MB respectively, the other datasets being roads and accidents data are small.

**DATASETS USED**

* Incidenti stradali: <https://opendata.comune.bologna.it/explore/dataset/incidenti_new/>
* Progetti città 30: <https://opendata.comune.bologna.it/explore/dataset/progetti-citta-30/>
* Quartieri di Bologna: <https://opendata.comune.bologna.it/explore/dataset/quartieri-di-bologna/>
* Velocita città 30: <https://opendata.comune.bologna.it/explore/dataset/velocita-citta-30/>
* Archi stradali: <https://opendata.comune.bologna.it/explore/dataset/rifter_arcstra_li/>
* Rilevazione flusso veicoli tramite spire - anno 2019: <https://opendata.comune.bologna.it/explore/dataset/rilevazione-autoveicoli-tramite-spire-anno-2019/>
* Rilevazione flusso veicoli tramite spire - anno 2020: <https://opendata.comune.bologna.it/explore/dataset/rilevazione-autoveicoli-tramite-spire-anno-2020/>
* Rilevazione flusso veicoli tramite spire - anno 2021: <https://opendata.comune.bologna.it/explore/dataset/rilevazione-autoveicoli-tramite-spire-anno-2021/>
* Rilevazione flusso veicoli tramite spire - anno 2022: <https://opendata.comune.bologna.it/explore/dataset/rilevazione-flusso-veicoli-tramite-spire-anno-2022/>
* Accuratezza flusso spire – 2019: <https://opendata.comune.bologna.it/explore/dataset/accuratezza-spire-anno-2019/>
* Accuratezza flusso spire – 2020: <https://opendata.comune.bologna.it/explore/dataset/accuratezza-spire-anno-2020/>
* Accuratezza flusso spire – 2021: <https://opendata.comune.bologna.it/explore/dataset/accuratezza-spire-anno-2021/>
* Accuratezza flusso spire – 2022: <https://opendata.comune.bologna.it/explore/dataset/accuratezza-spire-anno-2022/>
* Dati centraline qualità dell’aria 2017-2023: <https://opendata.comune.bologna.it/explore/dataset/dati-centraline-bologna-storico/>

**GITHUB REPOSITORY**

The repository can be found [here](https://github.com/marcomartini97/BTP).