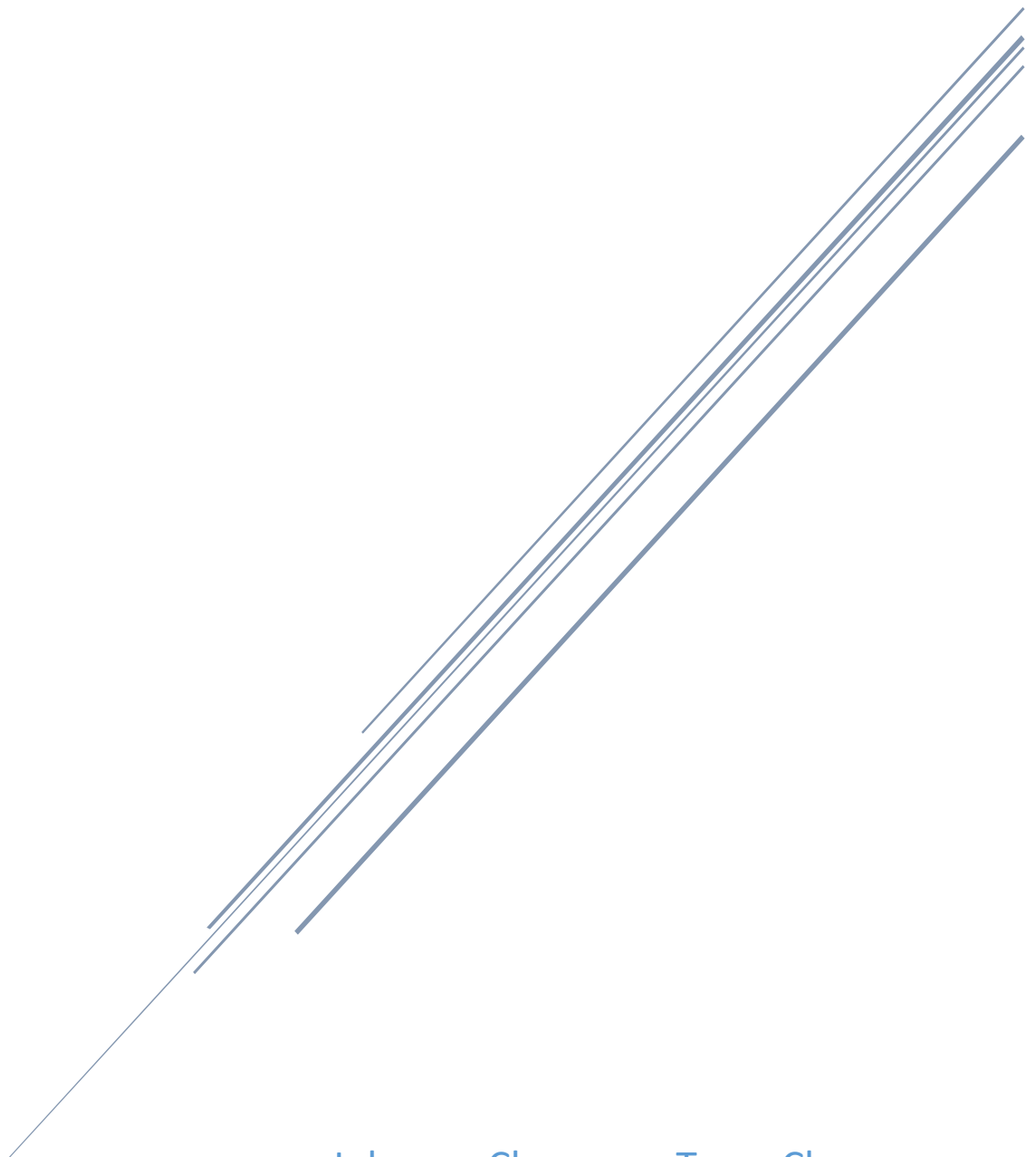


BUSINESS INTELLIGENCE PROJECT

Phase 1 report



Johanna Chapman, Tracy Chapman,
Clément Chaussade, Léo Fourier, Valentin Maréchal

Table of contents

Introduction.....	2
Work Organization	3
Additional data	3
Document description.....	4
Preliminary Workload.....	4
Dimensional Fact Model.....	4
Data dictionary	4
Additivity matrix	4
Formalization of requirements.....	5
Mapping table	5
Conclusion	5

Introduction

This report describes our work for the first phase of the Business Intelligence project. During this phase, we organized our team, analyzed the subject and the provided data sets, and established a first version of the data warehouse.

The objective of the project is to study the diseases of trees in the area of Grenoble. Three data sets have been provided: they describe various attributes of those trees; and while two of them describe their diseases, the last one doesn't. Our purpose is to study the first two data sets in order to predict the diseases of the trees in the last one.

Work Organization

For this phase, Johanna was the group leader. Regular meetings were organized in order to share our progress and distribute tasks.

The first objective of this phase was to understand the provided data and make some first analysis about the trees. Johanna, Clément and Valentin took care of it and created a first rough sketch of what the relational model could be. In the meantime, Tracy, Léo and Johanna started to seek for external data sources that could be added to our model.

After we achieved to create the relational model, all the group started to think about the Data Fact model we could use. The modeling was made by Johanna, Valentin and Léo. Johanna and Léo worked as well on the additivity matrix and the mapping tables. The preliminary workload and the formalization of requirements was done by Valentin. Clément produce the data dictionary. Meanwhile Tracy was in charge of placing the different trees we had on QGIS as well as the geographical external data sources.

Additional data

Several data sets found online have been considered to be added to our model. Data sets describing the city of Grenoble itself (buildings, roads) have been confirmed as usable and adequate. Pollution have been considered and are still being discussed as of now, so they are not present in the documents.

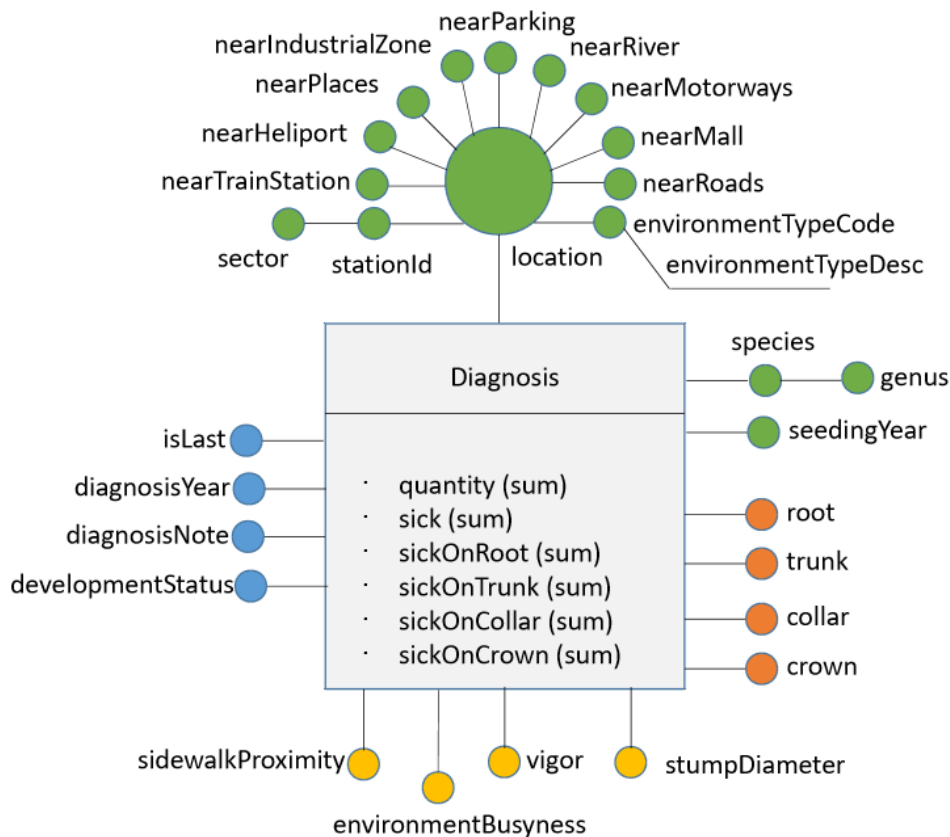
The data sources about the city were collected in [citer les sources] and the pollution was founded at [citer les sources aussi].

Document description

This section will describe each document produced during this phase. They can be found attached in the provided zip file.

Dimensional Fact Model

This diagram maps the Dimensional fact that has been designed to answer the user's needs, in accordance with the provided and additional data. The main fact is the diagnosis of a tree. Four categories of dimensions describe various attributes of the tree or the diagnosis: Location & Environment (green), Tree attributes (yellow), Diagnosis details (blue), and Diseases (orange).



Data dictionary

This table describes in further details the content of each attribute of the Dimensional Fact. For each attribute, it gives a precise description of its meaning, and the possible values it may have.

Additivity matrix

This table describes how measures are to be aggregated when combining dimensions.

Preliminary Workload

This document introduces eighteen queries that may be asked by the user in order to analyze the data. The data warehouse needs to be modeled in a way that provides quick and consistent answer to those queries, hence the importance of this document.

Formalization of requirements

This document associates each query defined in the Preliminary Workload with the dimensions and measures defined in the Data dictionary it requires the use of. It allows a quick see of the required measures and dimension for a specific query.

Mapping table

This final document describes where each attribute of the model has its data extracted from.

Conclusion

While phase 1 is over, decisions described above are still subject to change. External data sets are still being discussed and various documents will be updated in consequence.