Passport Project

Project for undergraduate students:

"IT services and technologies for processing data on transport"

Project mentor: Cutinel Maria Aleksandrovna  
Architect of implementations of HF Labs LLC

Project name: Analytics of data on quality metrics (with or without visualization)

The wording of the problem

Any MDM System (Master-Deninal Management) receives a large amount of data to the input.The data obtained must be correctly processed based on their specifics.To do this, first, it is necessary to conduct an initial analysis of data loaded from sources, identify errors, anomalies, features.

Also, after practicing the system algorithms, it is necessary to collect statistics on processed data in order to improve business processes

It is proposed to choose a set of statistical metrics for data analysis and visualize the results.

This will help:

Assess the quality of the input data, identify anomalies, erroneous values ​​in the dataset.

Assess the result of data processing by the system, adjust the settings depending on the quality of the input data.

Give recommendations to improve the quality of data in the source system.

Requirements for input competencies to participate in the project:

Be able to work with large volumes of structured data.

To be able to write SQL questions to the database.

Understand how the list of metrics is formed for the final visualization.

Be able to visualize aggregated data.

Starting data: synthetic data of individuals and legal entities.

Description of the result.The result that is planned to be obtained in the final of the project work with students:

Product: product that allows you to see the composition and quality of data and on request to form a report.

Educational:

experience in analysis of the subject area;

experience with large amounts of data;

experience in analyzing the quality of data and the formation of metric sets for the dataset;

Experience in visualizing the results on graphs and dashboard.

Student team composition: 2-5 people

Possible roles in the team:

Analyst;

Developer;

Tester;

UI/ux designer.

The number of student teams on the design task (if the competition is provided).

The schedule and working conditions on the product (stages and terms of the project, private tasks at each stage, tasks for various role subgroups).

Stages:

Description of the subject area of ​​the task and goals of the study;

Data analysis

The formation of a list of visualized metrics;

Selection of technological stacks;

Software development;

The formation of graphic elements of visualization;

Testing and debugging software;

Preparation of the final presentation/report with results, conclusions and recommendations.

The customer of the project (or to whom and which companies will be useful to the results of the project):

Customer: LLC Khf Labs.

Expert from the industry:

KUZERSTARYA Maria Aleksandrovna - architect of implementation projects of LLC Khf Labs

How the results of the project can be in demand in the industry (for the customer) - practical value:

The result of visualization of selected metrics will allow to evaluate the quality of the data loaded in the MDM system and the result of processing these data.

This will adjust the data processing algorithms, taking into account the data features, as well as form hypotheses to improve the quality of data coming from source systems.

Resource support for the project.To obtain the final product, the following according to:

Python programming language or similar;

Poe for data visualization, for example, Tableau Public, Redash or analogues.

BD to choose, for example, postgreSQL, clickhouse, sqlite;DB client (dbeaver and analogues)