

Documentation of the third phase of the data mining project

Faridreza Mumtazzandi 9812762601

Alireza Nurbaksh 9812762496

1 .Introduction:

From the knowledge that we gained with the datasets in the last two phases, we were able to reach a correct and comprehensive understanding of why and how the information we have and in the second phase to check the quality of the data, now finally in this phase we can ask comprehensive questions and It is important to answer regarding the dataset.

Because in this phase the project was divided into five different packages, by choosing the second package, we became responsible for analyzing and obtaining information related to the types of assets, which further examined each section and presented our solution about the solution. We will express these problems. In this phase, we will mainly deal with INOUT and INOUTLINE datasets because these two datasets are assigned to input assets.

2 .Statistics related to types of assets:

Our first mission in this phase is to obtain the rate of asset entry by year, which we took two ways:

- Review based on fiscal year:** Unfortunately, most of the records do not have exact values and most of them are null or non-numeric.

- Review based on the actual year:** as stated in the pdf file with the documentation, we use the actual year that the assets were registered in the dataset to fully understand the number of goods that were registered in the two datasets in each year. We will find and analyze based on it.

The points obtained from this analysis and the codes related to it can be seen completely in the pdf file of the project.

3 .Inspection of incoming goods based on deputy and holding company:

In this section, we want to examine both datasets about holdings and incoming goods. We will do this with the two attributes ACCT_AC_HOLDING_ID, which is only seen in the INOUT dataset, and the VAHED_MALI attribute, which is present in both datasets.

After using these two features, we can get interesting results and get a proper analysis of incoming goods and holding and accounting units responsible for assets. This information can also be seen in the main project file.

4 .List of properties that have different accounting document conditions:

In the final part of this phase, which can only be evaluated in the INOUT dataset, the registered goods are checked based on their accounting document status, and after obtaining each of the finalized groups, new datasets are drafted and awaiting approval based on the Obtained information and we will make different groups.

5 .Conclusion:

After the first two phases, in the second phase, we were able to easily obtain the right information and provide appropriate and complete answers, based on which we can make the final analysis and appropriate predictions in the final phase.