PWC Use cases
PROJECT II

Pauline Chaland, Romain Gangloff, Jason Maige, Vivek Pandya

CLUSTER ENTITIES AND ANALYZE EACH CLUSTER

PWC financial information for several entities have been provided to us for a specific financial year.

PROBLEM STEPS :

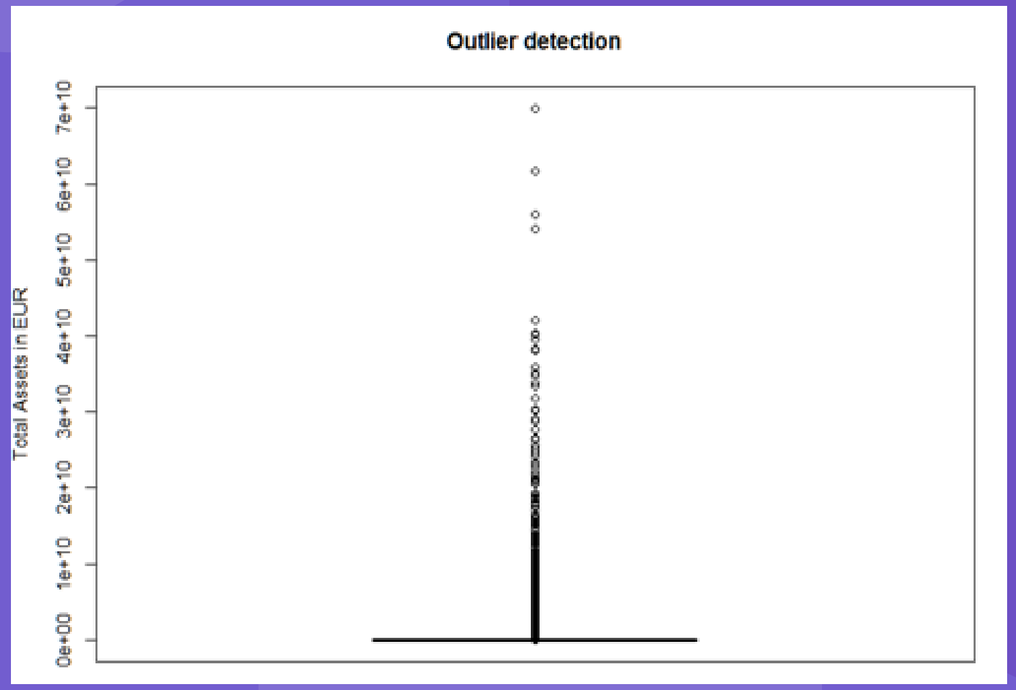
1. Data preparation & cleaning
2. Analysis of the features
3. Clustering Techniques
4. Visualization

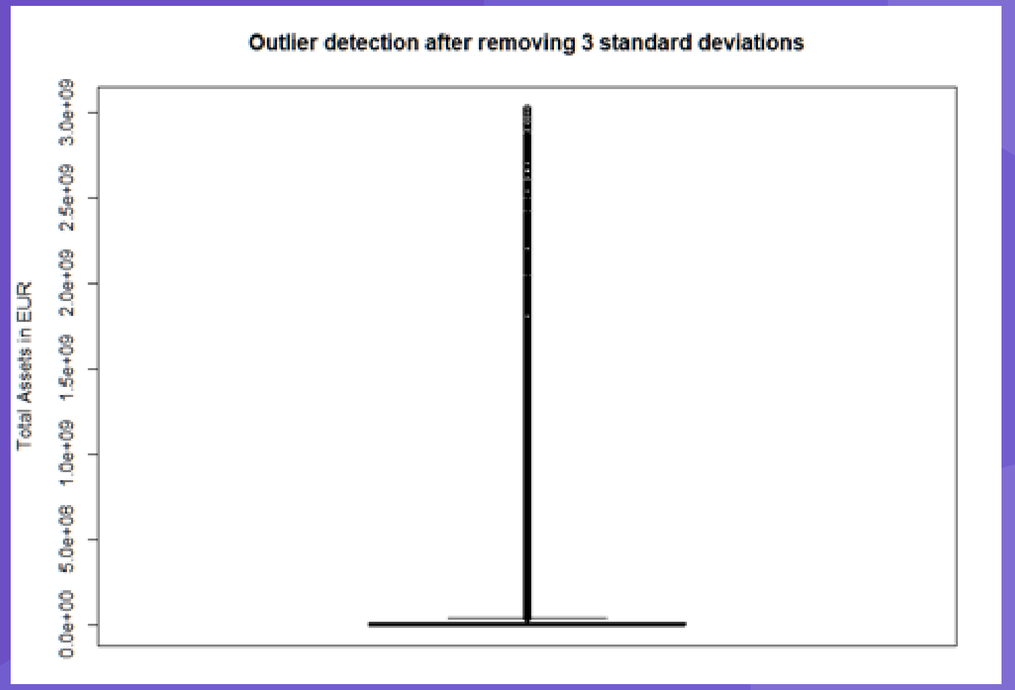
TABLE OF CONTENTS

1

DATA PREPARATION / CLEANING PROCESS

- Currencies convertion
- Basic statistical test : identification of extreme values
- Basic box plots : indentification of outliers
- Removing everything beyond 3 standard deviations
- From 72644 companies to 69047 companies



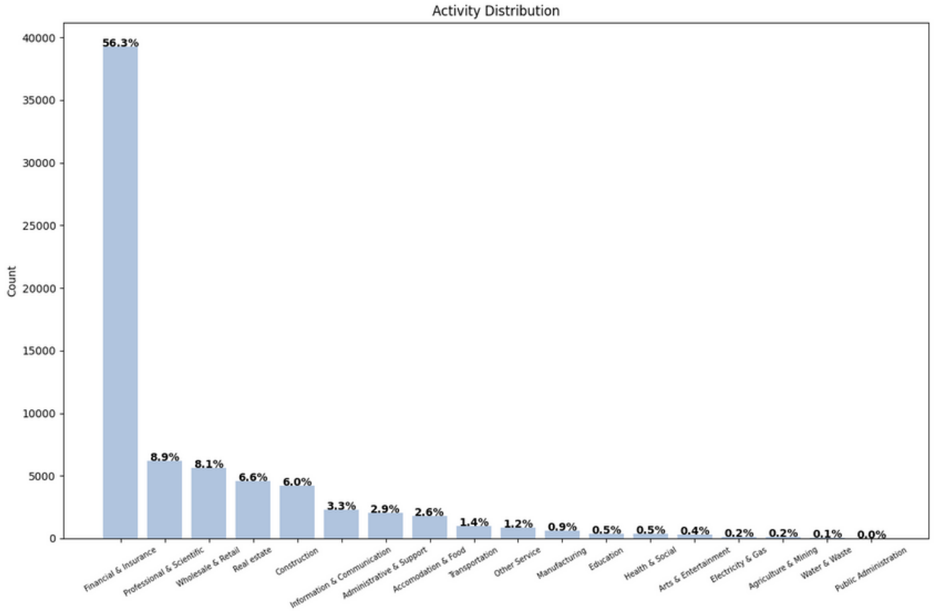


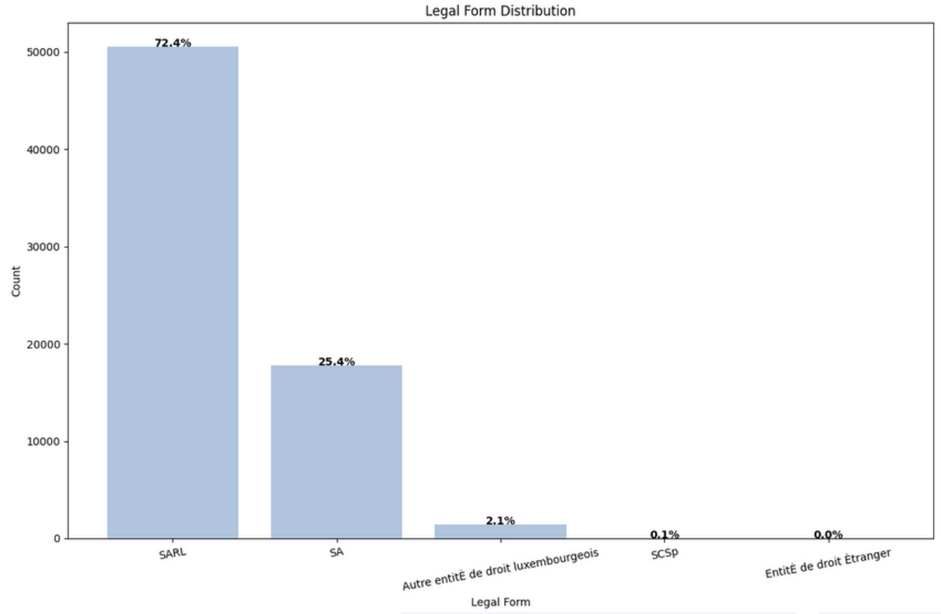
2

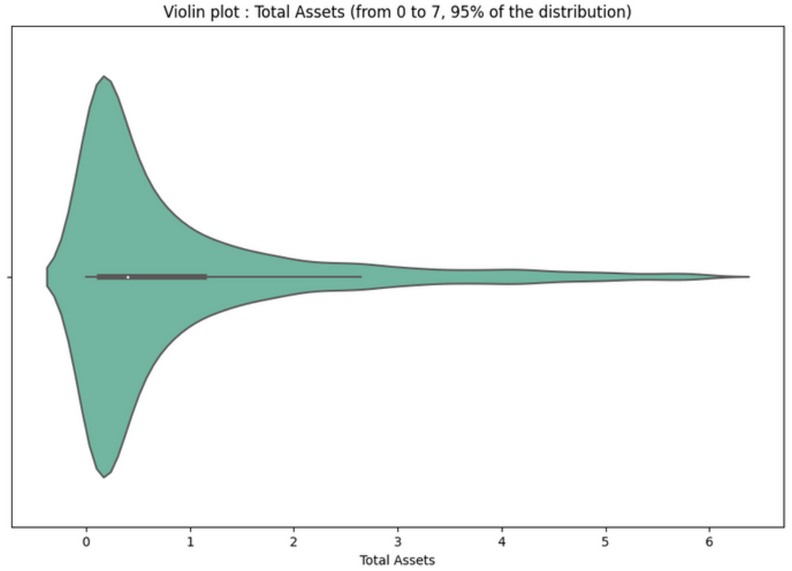
Summary statistics

DISTRIBUTION

- 56% in Finance & Insurance
- Massive representation of SARL (75%)

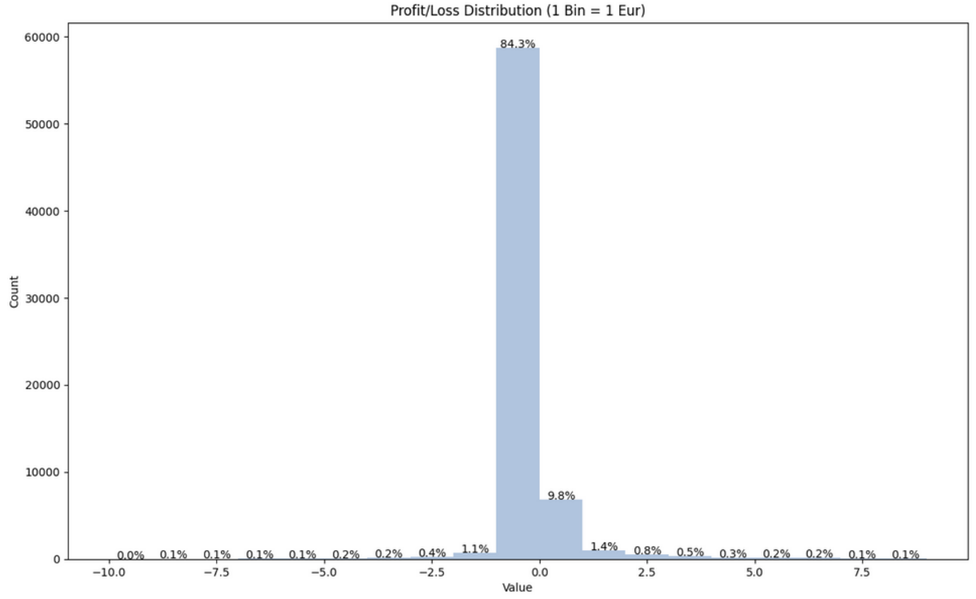






- Dominance of low value assets (between 0 and 1 eur).

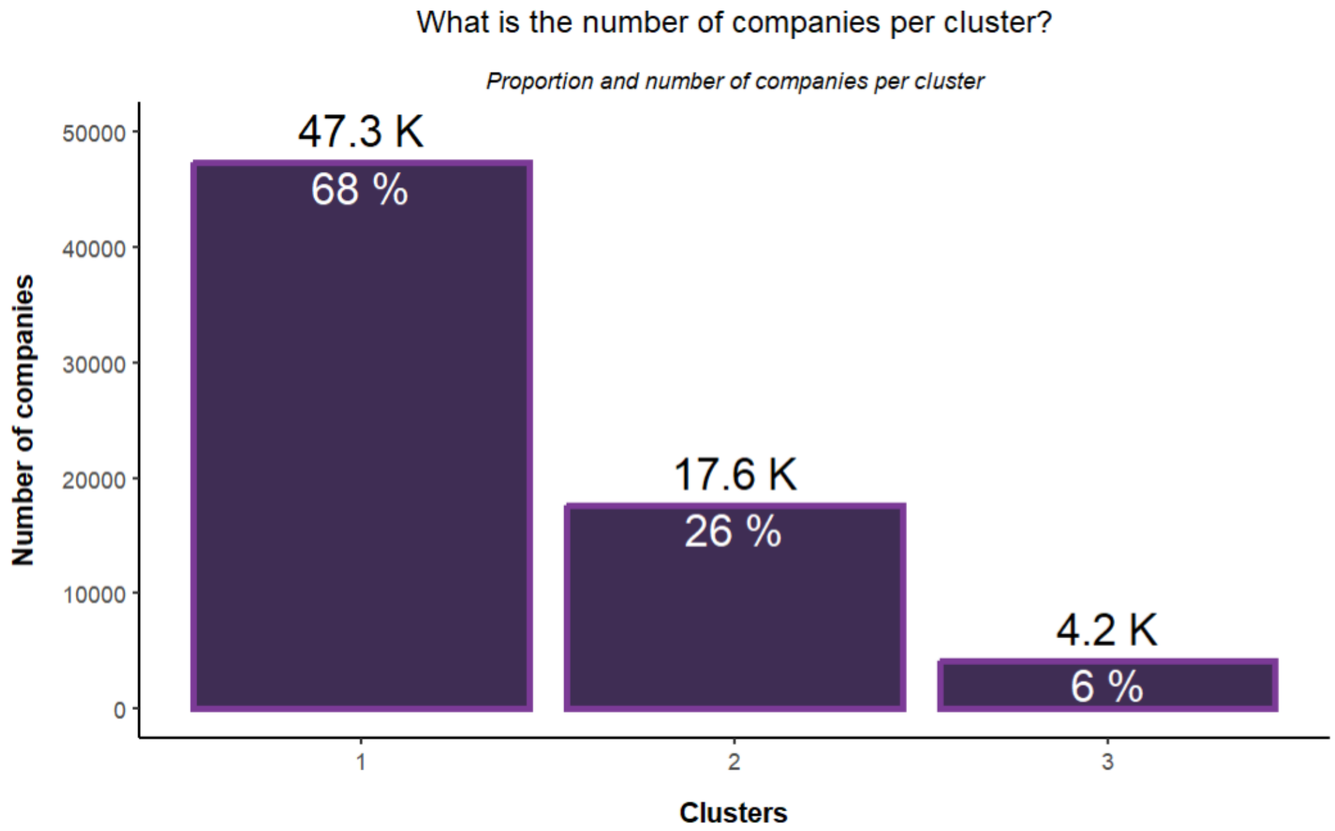
- Massive domination of a Profit/Lossbetween 0 and -1 (85%)
- So between -3M and 0,3M €



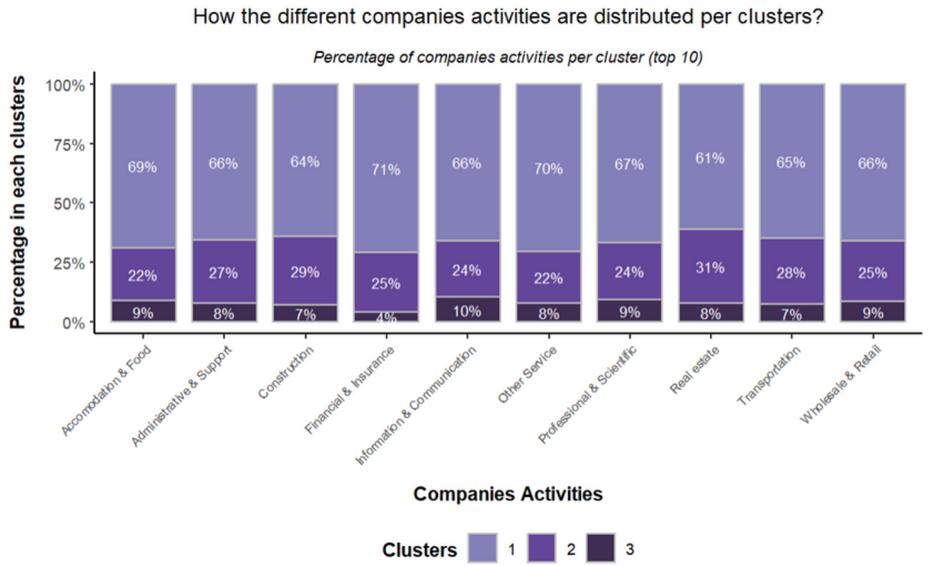
3 Presentation of the features

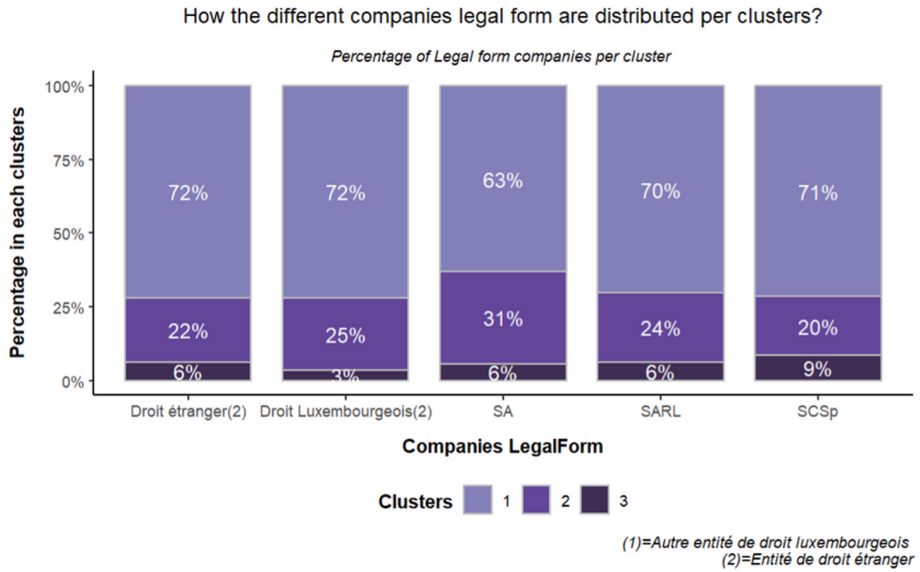
Clustering Techniques

- K mode algorithm
- Grouping profit and assets per group of 5
- Cluster 1 dominates the clustering



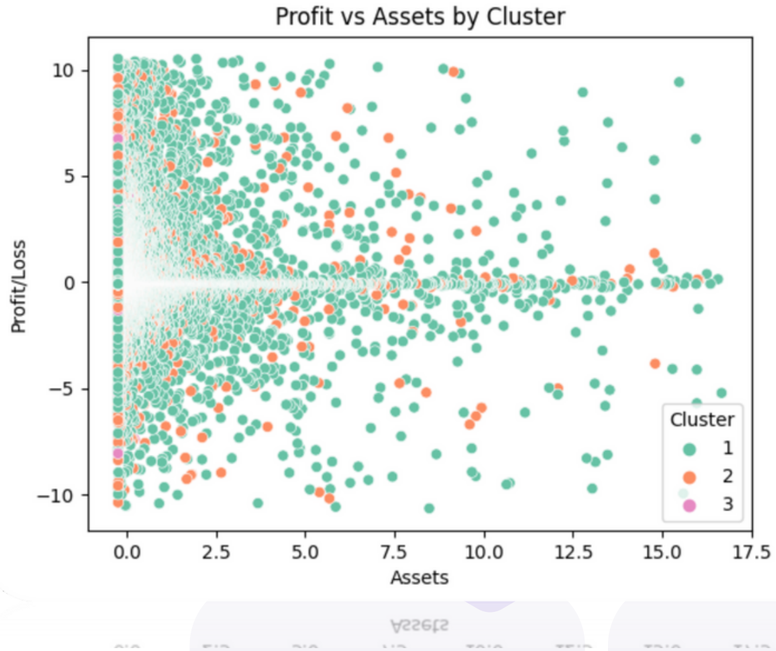
- Analysis of Company Legal form and Activities influences in the Clustering





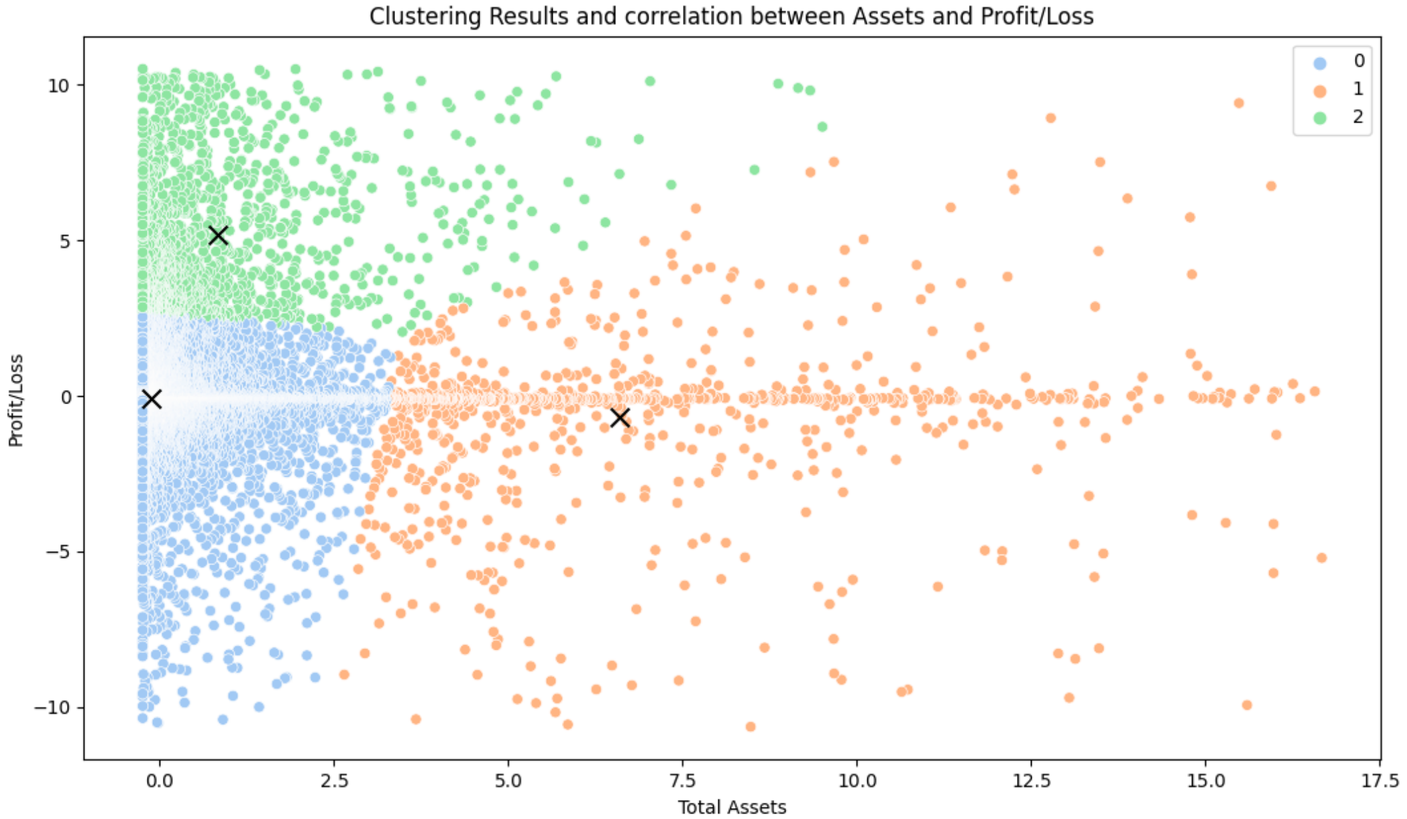
- No influences from Activities and Legal Form

Clustering visualization with Profit and Asset



- Slight influence from Profit and Asset

If Profit and Assets would have only been taking into account



4 Recommendations

Practical implementation

- The companies have similar characteristics
- No singularity
- Strong business base

- The clusters are mostly very dense.

THANKS!

Any questions?