

PE-BI REPORT

PHASE 0

Securities - Customer Profiling and Prediction

Team Members

Malek Douik

Amin Lazreg

Moez Abid

Chayma Ben Nacer

Koussay Ben Mhanna

Skander Marnissi

Class:

4DS2

Table of Contents

1 INTRODUCTION.....	3
2 PROJECT CONTEXT	3
3 DATA SOURCE IDENTIFICATION AND DESCRIPTION	4
4 BUSINESS OBJECTIVES	10
4.1 SCORING AND CLASSIFICATION	10
4.2 PREDICTION	10
5 SYSTEM ARCHITECHTURE DIAGRAMS	10
6 DATA WAREHOUSE SCHEMA	12
7 SOFTWARE ENVIRONEMENT	14
7.1 PYTHON.....	14
7.3 SQL SERVER DATA TOOLS	14
.....	14
7.4 SQL SERVER INTEGRATION SERVICES	15
.....	15
7.5 POWER BI	15
.....	15
8 REPORTING.....	16-20
9 MACHINE LEARNING	21-23
10 CONCLUSION.....	23

1 Introduction

Business intelligence (BI) includes the strategies and technologies used by enterprises for the analysis of data.

BI technologies provide historical, current and predictive views of the enterprises operations, common functions of business intelligence technologies include reporting, analytics, data mining, text mining, predictive analytics and prescriptive analytics.

Thus, BI offers a huge variety of tools, methodologies and applications to collect data from internal data base and external sources, transform them to make analysis easier, load them and launch queries.

These tools can then be used to create reports, dashboards, and data visualizations to facilitate the interpretation to the manager.

2 Project context

ODDO BHF is an independent Franco-German financial services group, with a history stretching back over 150 years. With 2,300 employees (1,300 in Germany and 1,000 in France), and more than 100 billion euros in assets under management, ODDO BHF operates in three main businesses, based on significant investment in market expertise: private banking, asset management and corporate and investment banking.

ODDO BHF investment banking referred to us to participate in putting in place a prediction, profiling and classification of clients/prospects of ODDO BHF to better understand their financial behavior and identify the hidden opportunities.

Moreover, ODDO BHF aim to predict opportunities and commercial risks like sales growth and sales diminution to optimize the client service and to guarantee its satisfaction.

ODDO BHF is proposing to use their internal data like its client's transactions and interactions from their CRM, and external data from social networks, and other financial and economical source.

3 Data Source identification and description

CRM_V_CONTACTS. Csv :

Repository of prospects/ customers known as "contacts" with whom ODDO BHF has a commercial relationship.

AccountId	Unique account identifier in the CRM to which the contact belongs
ContactId	Unique technical ID
Function	Function Label
IssuerFunction	Function Label
Job	Wording profession
Ville	City
CreatedOn	Technical creation date
UpdatedOn	Date of last technical update
BE_EQ	Subscription to the financial analysis: Equities
BE_FI	Subscription to the financial analysis: Fixed Income
BE_ECO	Subscription to the financial analysis: Economique
BE_ESG	Subscription to the financial analysis: ESG
BE_CHR	Subscription to technical financial analysis

PMU_RCPT_COMPTE.csv :

Reference scale for the accounts of Securities clients.

CLT_IDCLT	Unique identifier of the customer to whom the account is linked
GCO_CODCOMPTEGCO	Unique account identifier
CPT_INDACTIF	Account activity indicator (0 inactive / 1 active)
CPT_DATOUVERTURE	Account opening date
CPT_DATFERMETURE	Account closure date

CLI_GTI_GeneriquesTiers. Csv :

Reference table of third parties in relation with the bank (customers, counterparties).

GTI_CodSociete	Allows to know the business entity of the third party
GTI_CodPrestataire	Defines which back office system communicated this third party
GTI_CodTiers	Unique identifier of the third party
GTI_CodPaysNaissance	Country of birth code
GTI_AdrDepartement	
Naissance	
GTI_NumResidentFiscalite	
GTI_CodCivile	Code of politeness

GTI_NumStatutPers	Third party type (SAMIC opening code)
GTI_DatNaissanceOu Creation	
GTI_CodSexe	
GTI_CodSexe	
GTI_NumSituationFa miliale	Professional category code
GTI_CodCSP	
GTI_NumProfession	Type of investor
GTI_CodTypeInvestis seurMif2	Level of customer experience in finance
GTI_CodTypeInvestis seurMif2	Customer's ability to master the so-called complex products
Account Custody	

CLI_TCL_TiersComptesLocal.csv :

Links between a securities account (portfolio) and a third party.

TCL_CodCompte	Title Account ID
TCL_CodTiers	Third party identifier
TCL_NumLien	Third Party Attachment Code
CRT_Libelle	Lets you know the business entity of the account

CLI_GCO_GeneriquesComptes.csv :

Portfolio / Contract Reference Table.

GCO_CodSociete	Lets you know the business entity of the account
GCO_CodPrestataire	Indicates the source of the account)
GCO_CodCompte	Unique identifier of the title account
GCO_CodTypeOrientation	Free Axis of Account Categorization
GCO_IsOuvert	Indicates if the account is open
GCO_IsAsv	Indicates whether the title account is an asset sub-account or an insurer's account
GCO_IsAssureur	Indicates if this is a life insurance account
GCO_IsPea	Indicates if it is a PEA)
GCO_IsPeaPme	If Account title: Indicates if it is a PEA / PME)
GCO_CodProduit	Life insurance: product code
GCO_LibProduit	Product name
GCO_TypGestion	Type of mandate of the portfolio (Managed, free, ...))
GCO_LibEtatCompte	Calculated area to display the state of the account (blocked, rich, It is necessary to use rather the codes than this zone
GCO_CodTarif	Brokerage Rate Code
GCO_CodTarifDDG	Fee Code Custody

GCO_CodTarifFHG	Fee Code Management
GCO_IsNanti	Indicates if the account is pledged

CRO_CRO_CROD.csv :

Detailed operations report(SAMIC, SAB,...)

CRO_CodSociete	Lets you know the business entity of the account
CRO_CodCompte	Unique account identifier
CRO_Dateffet	Effective date of transaction on account positions
CRO_CodOperation	Code Operation (procedure)
CRO_LibOperation	Operation label (linked to CodTypOperation)
CRO_CodAnnulation	Cancellation Code (A = Canceled, N Canceling))
CRO_CodSens	Code meaning of the operation
CRO_CodIsin	Isin code identifying the instrument
CRO_Qte	Decimated quantity
CRO_Crs	Trading course
CRO_MntBrutDevDe p	Gross amount of the transaction expressed in the currency of the vote
ESO_CodType	(Identify the type of operation: F -> Species T -> Titles E -> Exotic Performance)

ESO_CodProvenance	Identifies the source of funds: E - External to Institution I - Internal to Institution)
-------------------	--

INS_GIN_GeneriqueInstruments.csv

Table of financial instruments that may be held (in position) by clients.

GIN_CodISIN	Lets you know the business entity of the account
KW02_Libelle	Unique account identifier
GIN_Actif	Effective date of transaction on account positions
GIN_CodMIC	Code Operation (procedure)
GIN_CodPaysEmission	Operation label (linked to CodTypOperation)
GIN_CodProduitComplexe	Cancellation Code (A = Canceled, N Canceling))
GIN_LibInstrument	Instrument label
GIN_LibEmetteur	Instrument transmitter label
GIN_CodDeviseCotation	Code de la devise de cotation

Data Source identification and description

4 Business objectives

a. Scoring and classification

i. Client Segmentation

Grouping customers according to their financial movements and commercial aspects (Industry, localization, net worth).

ii. Monitoring operations, accounts, contacts and contracts

Monitoring operations details and grouping them by client and by period.

Grouping accounts by region and by period and sorting them by brut amount.

Grouping contacts by type and sorting them by the highest activity.

Monitoring contracts details and grouping them by type, amount ...

General monitoring of investments

iii. Web scraping

Create a web scraping in order to get the impressions and opinions of internet users on ODDO BHF.

b. Prediction

i. Classify the clients according to clusters detected in the data.

ii. Predict a client brut amount from multiple variables.

5 System architecture diagrams

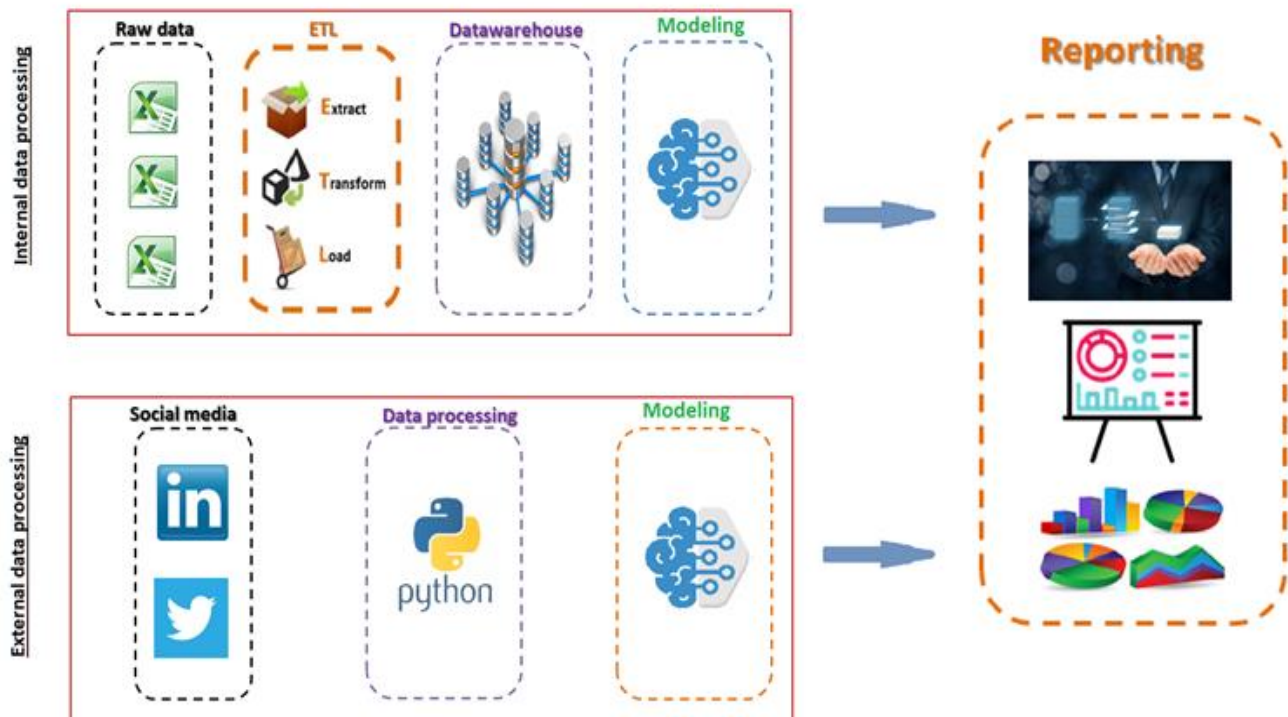


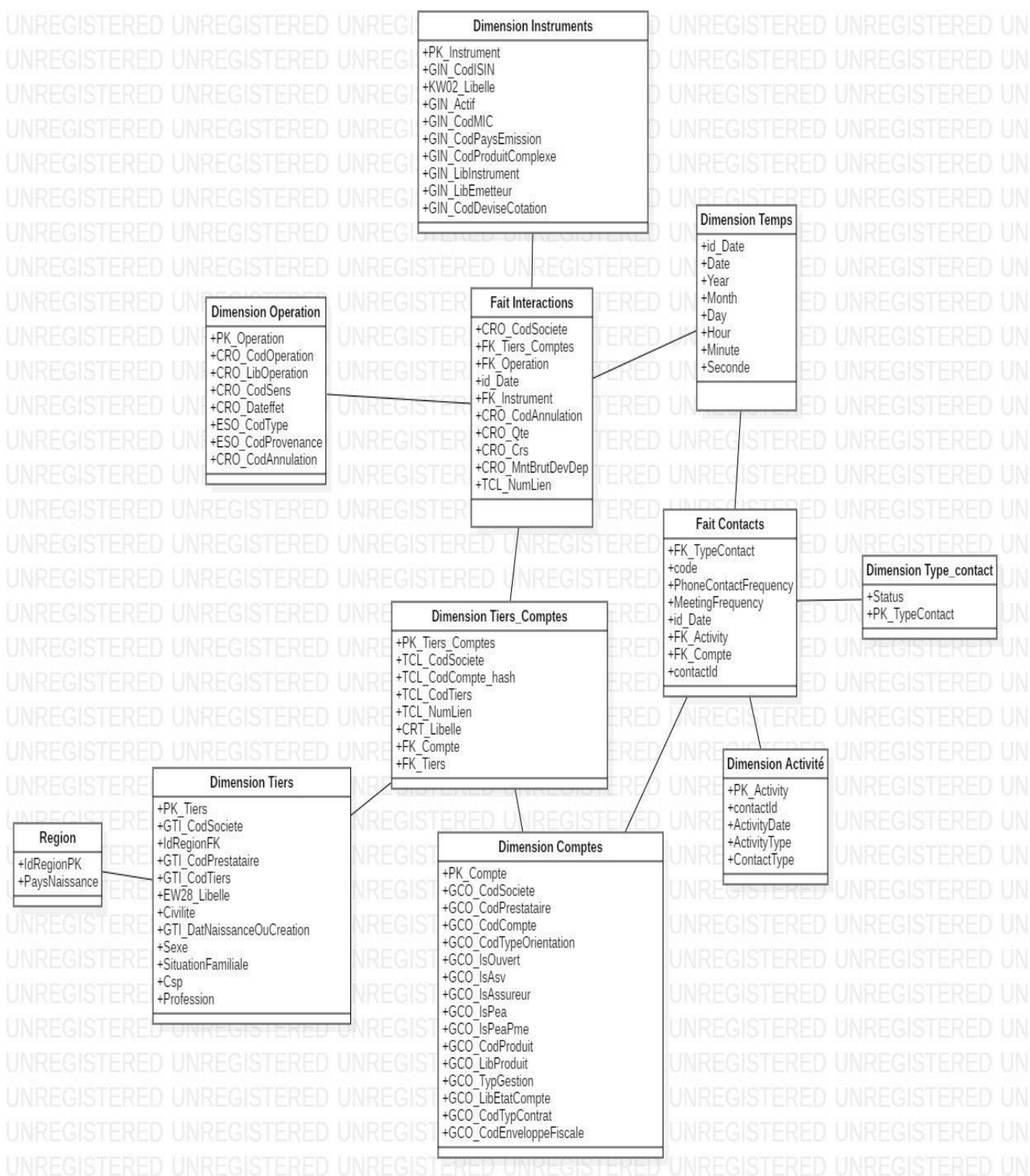
Figure 1: Architecture diagram

Fact table description :

- **Fait Interactions** : Details all the operations made by the client's financial account.
- **Fait Contacts** : Details all the contacts made between the client and ODDO BHF.

Dimensions table description :

- **Dimension Instruments** : Describes all types of financial assets such as stocks, credits ...
- **Dimension Temps** : Contains the time factor of every Fact table.
- **Dimension Operation** : Details all the operations made by the client's financial account.
- **Dimension Type_Contact** : Contains the interactions types.
- **Dimension Tiers_Comptes** : Represents the intermediary between Dimension Tiers and Dimension Comptes.
- **Dimension Activité** : Contains the interactions.
- **Dimension Comptes** : Describes the financial account.
- **Dimension Tiers** : Describes the client's information.
- **Region** : Represents the country of the client.



7 Software environment

7.1 Python



Figure 2: Python

Python: is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. It allows you to create data structures that can be re-used, which reduces the amount of repetitive work that you will need to do.

It is a powerful programming language used for many different types of applications within the development community. Many know it as a flexible language that can handle just about any task.

7.2 SQL Server Data Tools



Figure 3: SQL Server Data Tools

SQL Server Data Tools : is a modern development tool for building SQL Server relational

databases, Azure SQL databases, Analysis Services (AS) data models, Integration Services (IS) packages, and Reporting Services (RS) reports.

7.3 SQL Server Integration Services



Figure 4: SQL Server Integration Services

SQL Server Integration Services (SSIS) : is a component of the Microsoft SQL Server database software that can be used to perform a broad range of data migration tasks.

SSIS is a platform for data integration and workflow applications. It features a data warehousing tool used for data extraction, transformation, and loading (ETL). The tool may also be used to automate maintenance of SQL Server databases and updates to multidimensional cube data.

7.4 Power BI



Figure 5: Power BI

Power BI: is a business analytics service by Microsoft. It aims to provide interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards.

8 Reporting

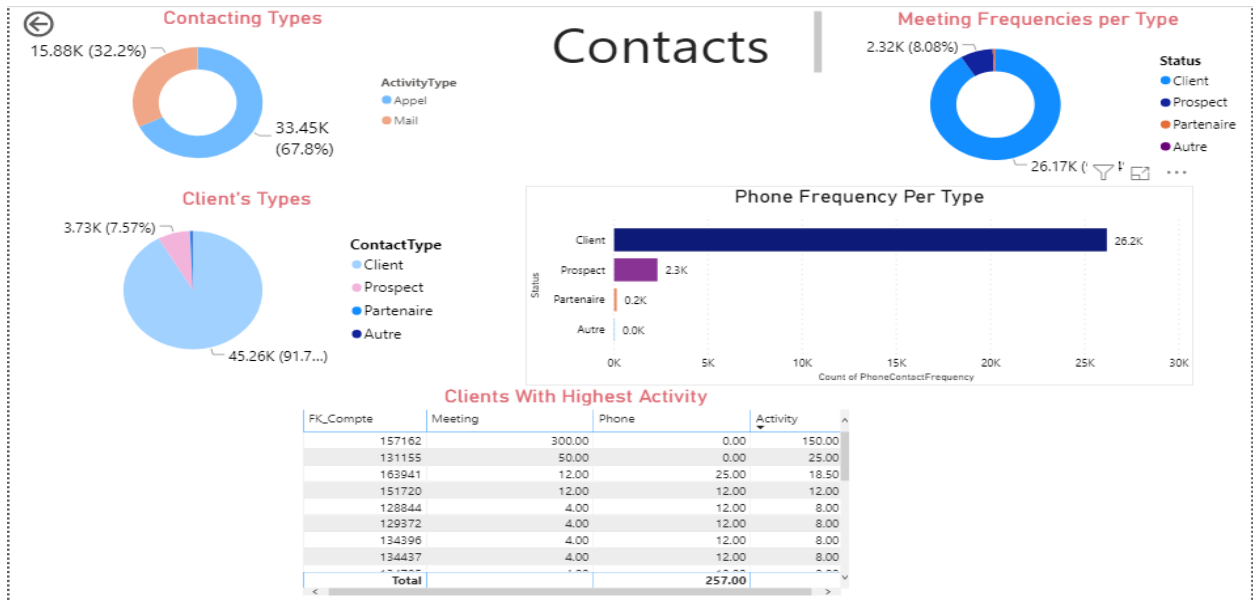


Figure 6: Contacts dashboard

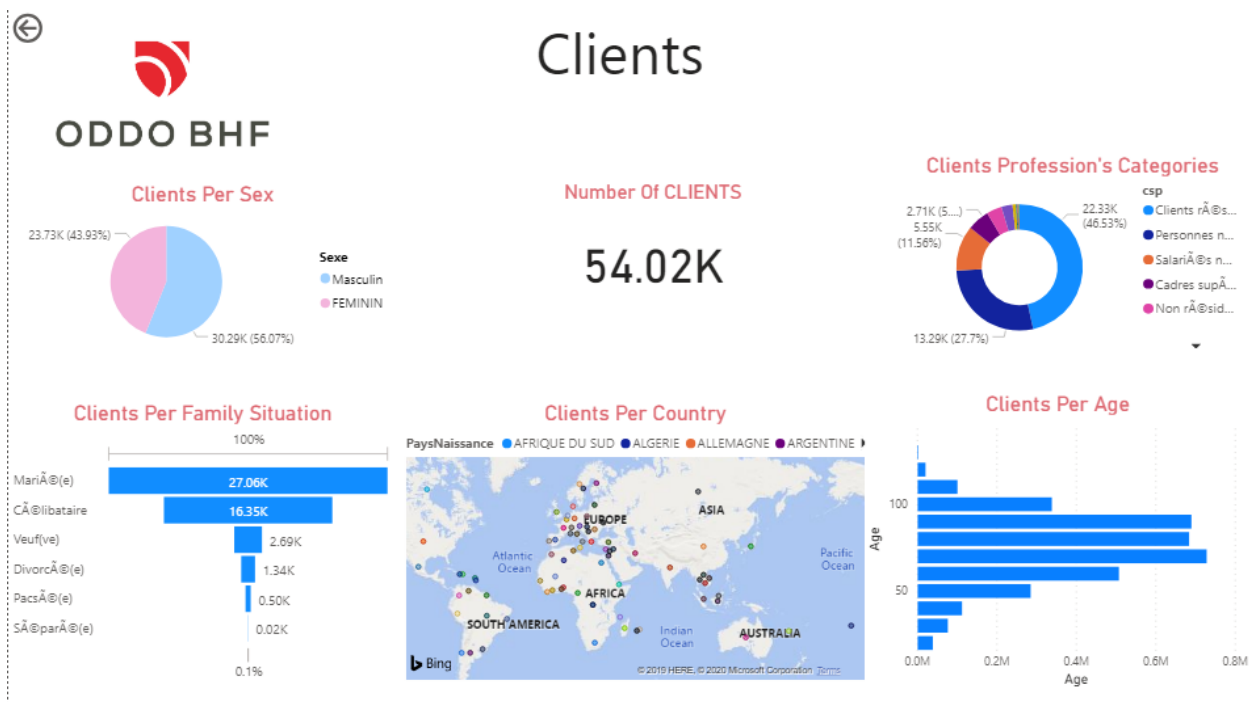
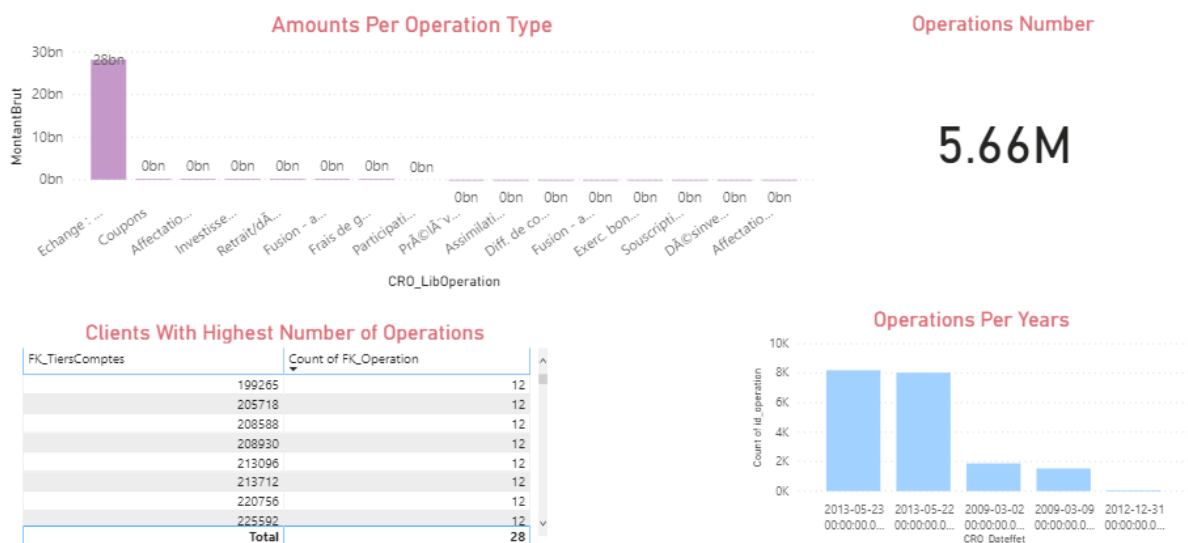


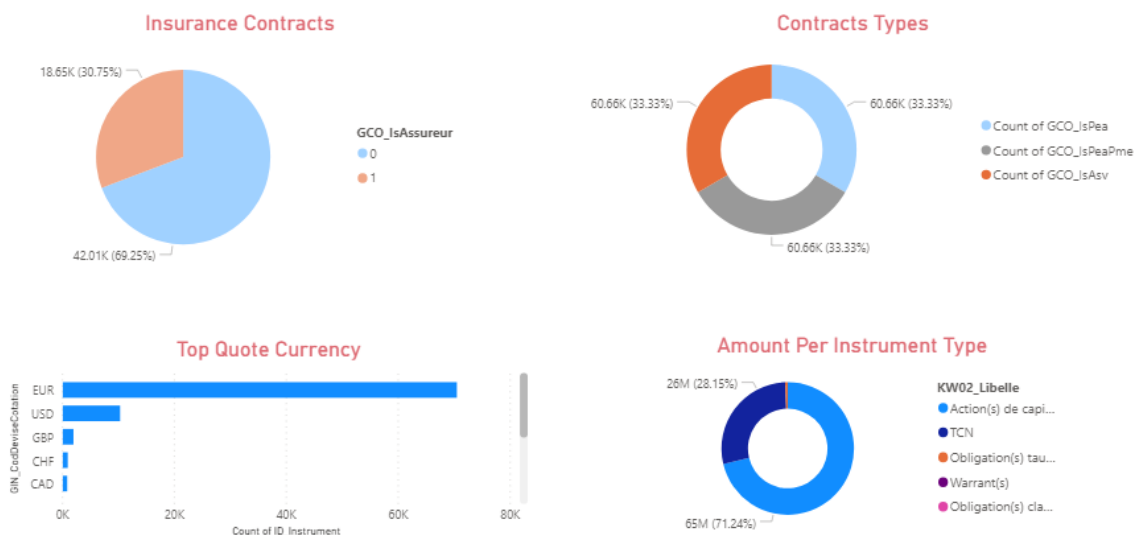
Figure 7: Clients dashboard



Operations



Contracts And Instruments



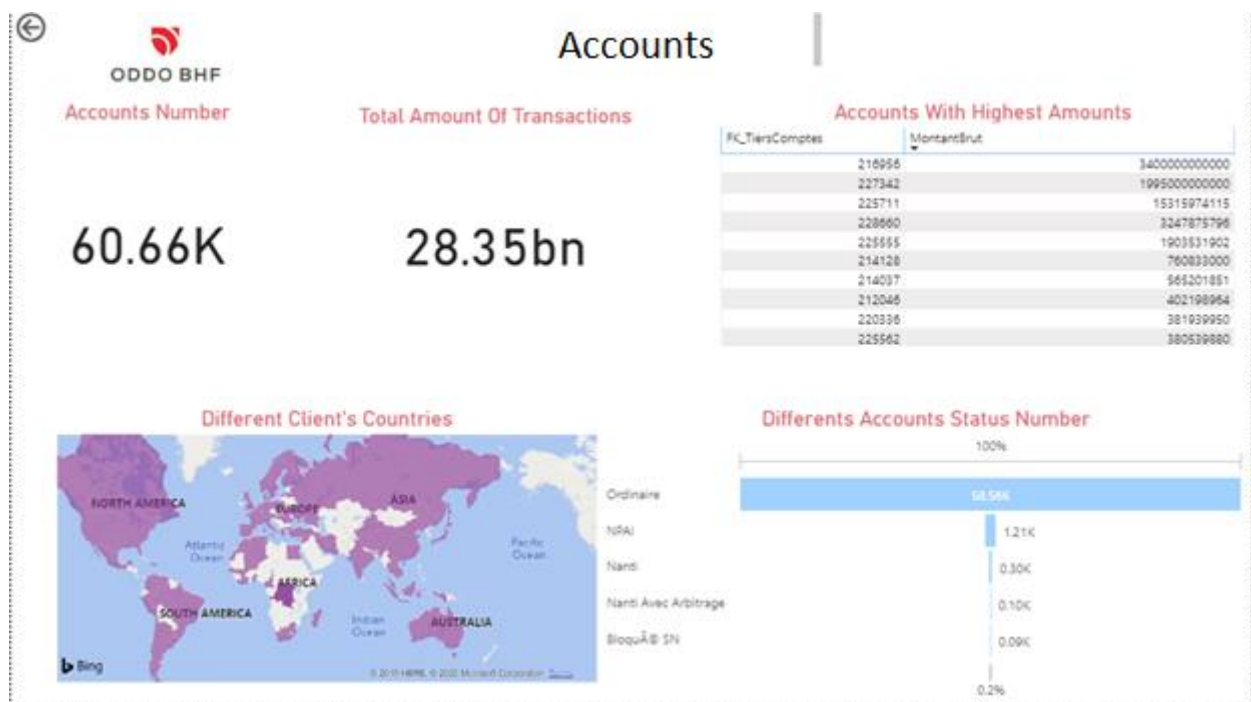


Figure 10: Accounts dashboard

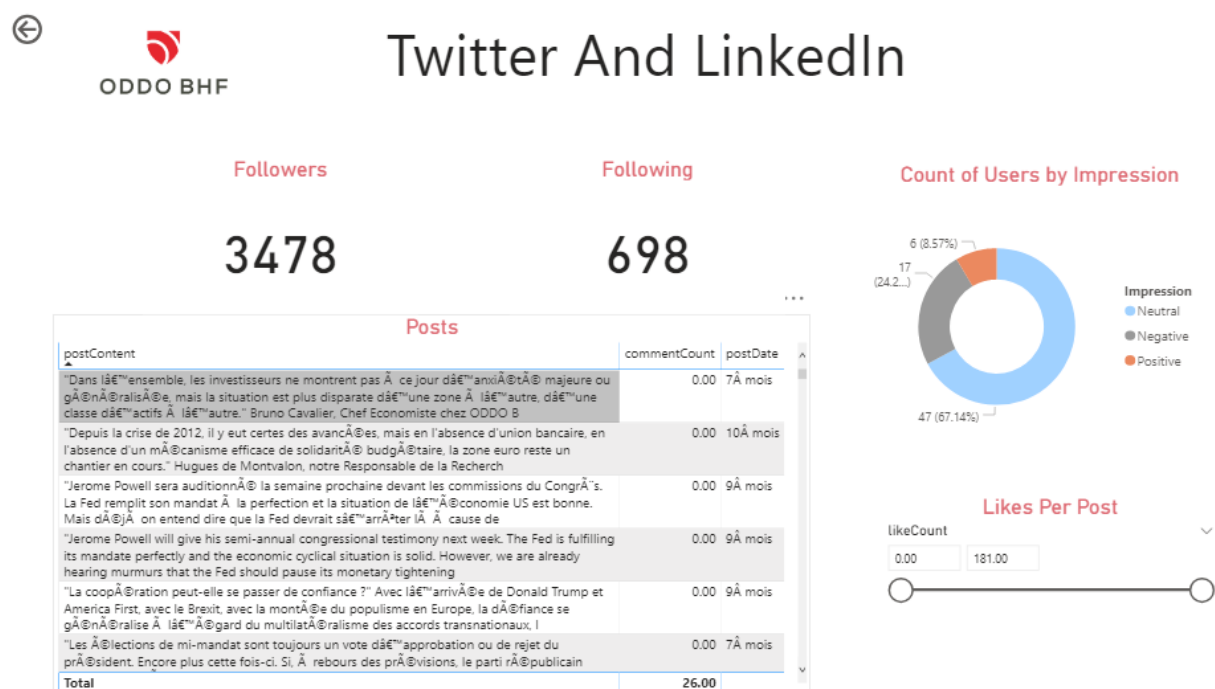
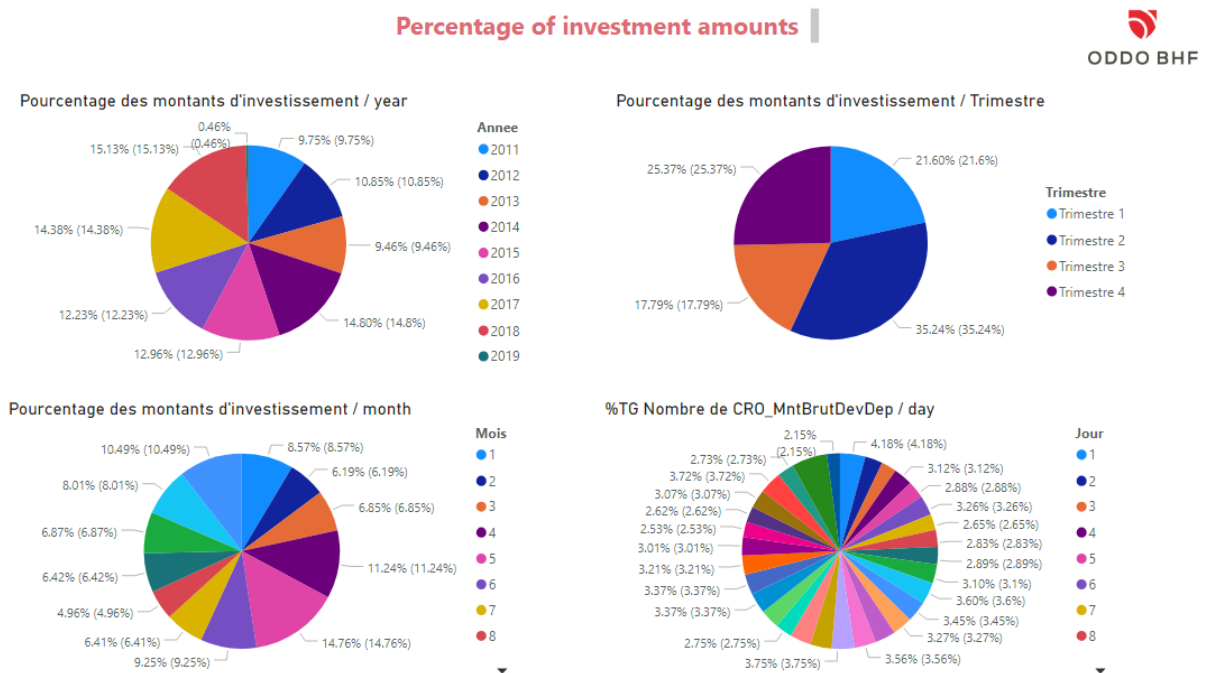
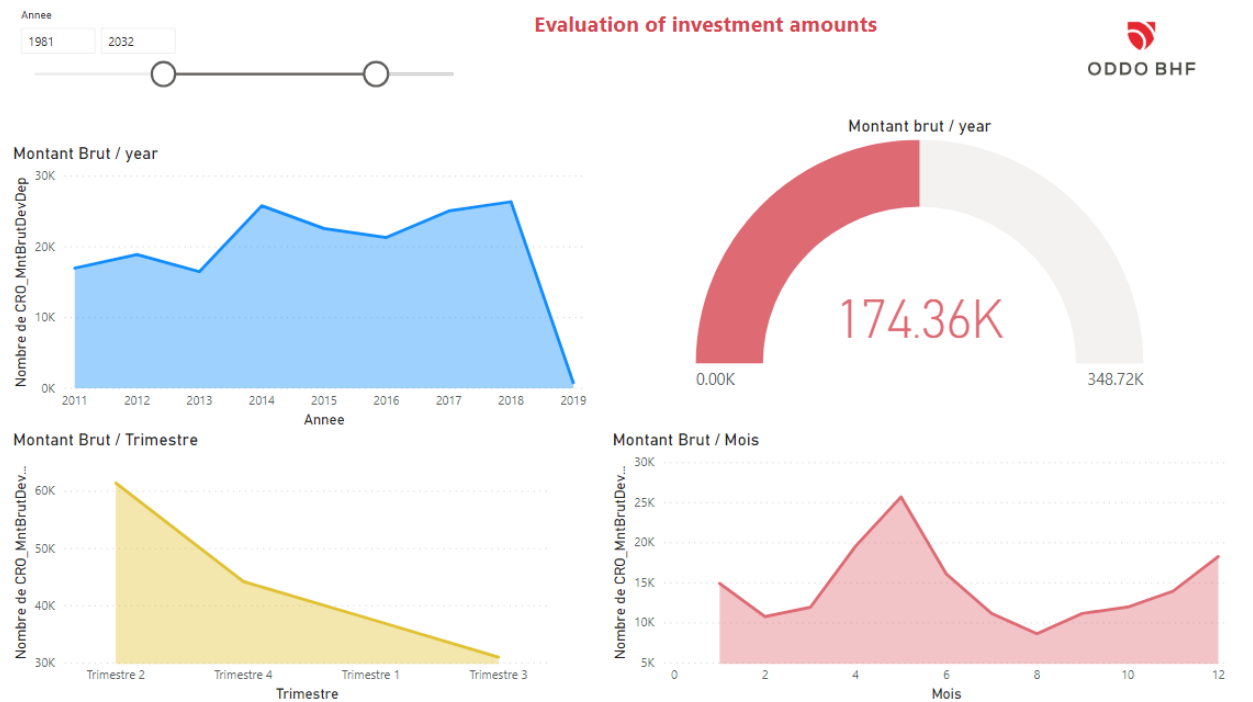


Figure 11: Social media dashboard



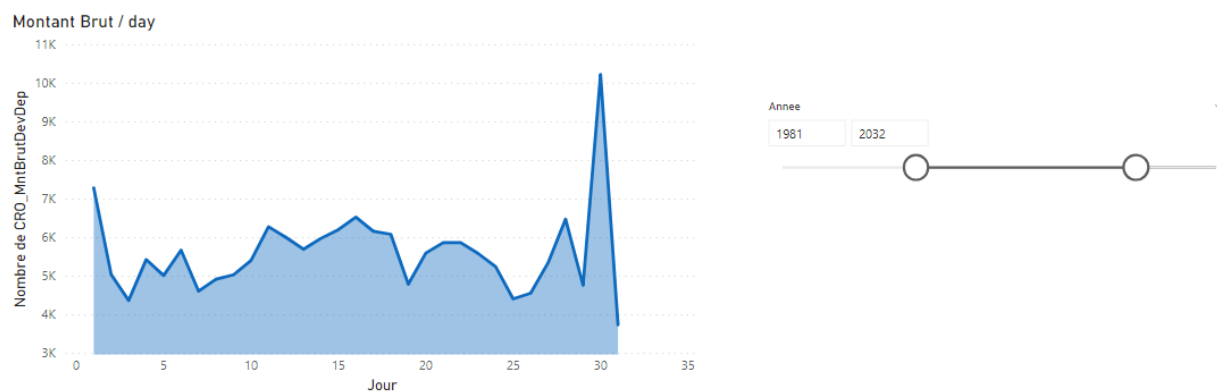


Figure 14: Brut amount

9 Machine learning

Cross-validation

```
def evaluate(algo, X_train, Y_train, cv):
    # One Pass
    model = algo.fit(X_train, Y_train)
    acc = round(model.score(X_train, Y_train) * 100, 2)

    # Cross Validation
    train_pred = model_selection.cross_val_predict(algo,
                                                    X_train,
                                                    Y_train,
                                                    cv=cv,
                                                    n_jobs = -1)

    # Cross-validation accuracy metric
    acc_cv = round(metrics.accuracy_score(Y_train, train_pred) * 100, 2)

    return train_pred, acc, acc_cv

def fit_ml_algo(algo, X_train, y_train, cv):
    # One Pass
    model = algo.fit(X_train, y_train)
    acc = round(model.score(X_train, y_train) * 100, 2)

    # Cross Validation
    train_pred = model_selection.cross_val_predict(algo,
                                                    X_train,
                                                    Y_train,
                                                    cv=cv,
                                                    n_jobs = -1)

    # Cross-validation accuracy metric
    acc_cv = round(metrics.accuracy_score(Y_train, train_pred) * 100, 2)

    print("Model used :", algo.best_estimator_)
    return train_pred, acc, acc_cv
```

Figure 15: Cross validation code

Hierarchical clustering

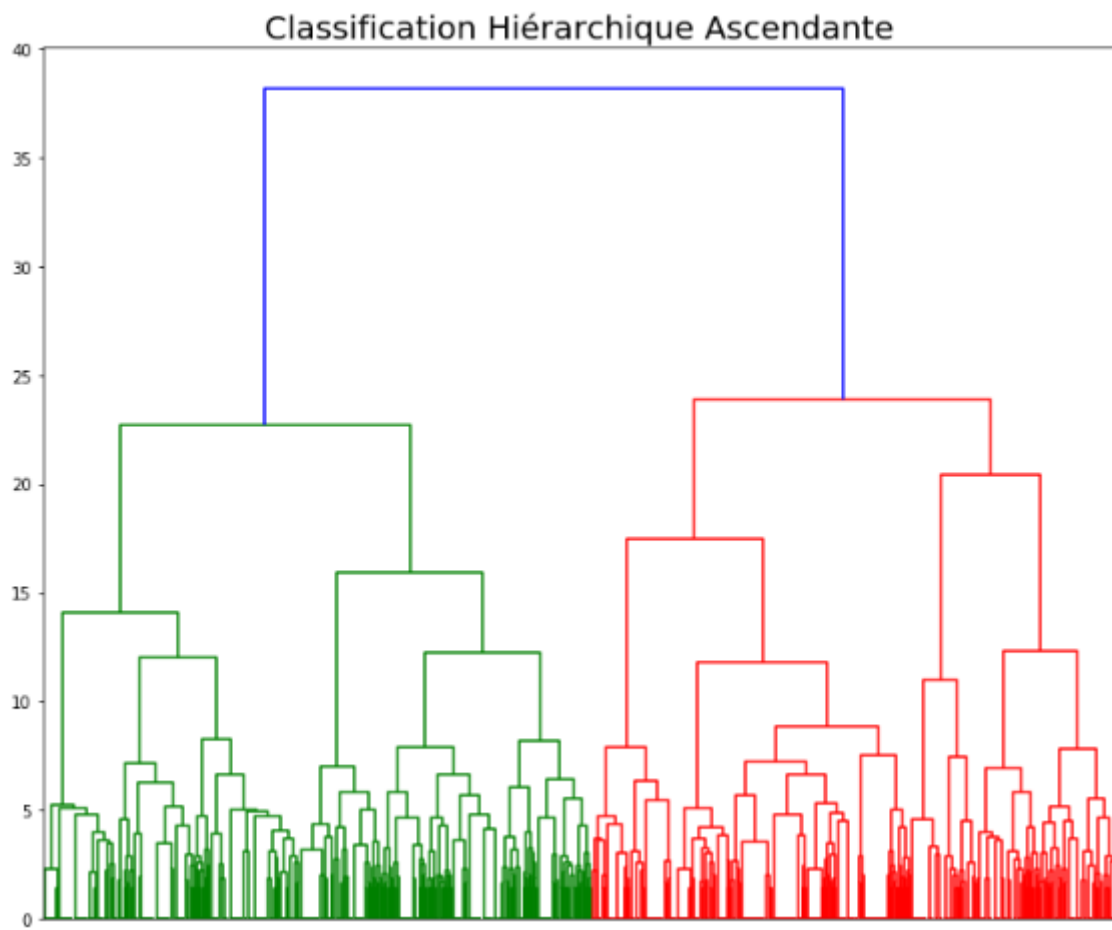


Figure 16: HC output

Multiple regression

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)
```

```
X_test.reshape(-1,1)
```

```
array([[ -0.13351391],  
       [  0.95671325],  
       [  0.          ],  
       ...,  
       [  1.3021721 ],  
       [-0.35675303],  
       [-1.03279556]])
```

```
y_pred = linear_regression.predict(X_test)  
y_pred.shape
```

```
(186,)
```

```
y_test = np.array(y_test)  
y_test.shape
```

```
(186,)
```

```
from sklearn import metrics
```

```
# say you have a trained model, clf
```

```
print("Le score du modèle de régression multi-linéaire est : ",abs(linear_regression.score(X_test, y_test).round(2)))
```

```
Le score du modèle de régression multi-linéaire est :  4.6512542317223993e-27
```

Figure 17: Multiple regression code

10 Conclusion

To deliver a Business Intelligence solution, we have gone through understanding the provided data, ETL technics and elaborating a data warehouse. Next, in order to enhance the decision making process, we were able to extract data from the web. After that, we started analyzing all of these data in order to achieve the business objectives. And finally, we were successful elaborated a dashboard for the reporting process and applied some artificial intelligence technics in order to have models with efficient results.