MASTER

MASTER'S IN DATA ANALYTICS FOR BUSINESS

MASTER´S FINAL WORK

PROJECT

BUILDING A BUSINESS INTELLIGENCE MODEL FOR "STEELOUETTE"

GERARD NAJEM

MARCH - 2023

MASTER

MASTER'S IN DATA ANALYTICS FOR BUSINESS

MASTER´S FINAL WORK

PROJECT

BUILDING A BUSINESS INTELLIGENCE MODEL FOR "STEELOUETTE"

GERARD NAJEM

SUPERVISION:

JESUALDO FERNANDES

MARCH - 2023

GLOSSARY

DDL – Data Definition Language

ETL – Extract, Transform, and Load

FK – Foreign Key

MFW – Master's Final Work.

PK – Primary Key

SQL – Structured Query Language

ABSTRACT AND KEYWORDS

This project aims to create a full business intelligence process for the company "Steelouette" based in Lebanon. First, by creating a data model, to then integrate their data in it, and use it for creating dashboards that will give them insights and information regarding their financial numbers. This project would be considered as an asset for the company and could be developed to contain broader data.

A conceptual and logical model were built based on discussions with the firm to understand their needs and the available data, then integrated it in the physical database, using a combination of 2 tools: Python and PostgreSQL. Later, dashboards have been created to highlight the main revenue numbers of the company using Power BI.

The results from the project are going to be used most importantly to gather insights about the business and maximize revenue and eventually profit, but also to highlight the gaps in the data entry that might have an impact on it.

KEYWORDS: Business Intelligence; Dashboards; Data; ETL; Modelling.

TABLE OF CONTENTS

TABLE OF FIGURES

Acknowledgments

I would like to express my deepest gratitude to my family and friends for their unwavering support and encouragement throughout my academic journey. Their love and guidance have been instrumental in helping me achieve my goals. I would also like to thank Professor Jesualdo Fernandes for their invaluable mentorship and expertise, which have pushed me to give my best even when I thought I could not. And without a doubt I would like to thank the company “Steelouette” and its co-founders Elsa Bakhos, and Ghadi Barhouche for their trust in my project and their support all along the journey. Without their guidance, I would not have been able to reach the level of academic success that I have today.

1. INTRODUCTION

In today's data-driven business environment, companies generate an enormous amount of data that requires efficient storage, management, and analysis to obtain valuable insights that can guide decision-making. In the modern business landscape, there is an unrelenting focus on the collection and organization of client information, social media analytics, and any other data that could be of value (Li et al., 2018; Kudyba & Hoptroff, 2017; Noori et al., 2020; Fan et al., 2014; Cukier, 2013). The strategic tracking of these metrics on a yearly, quarterly, or even daily basis is proving to be a powerful tool for businesses, as it can save them millions of dollars by allowing them to predict potential product failures or identify departments that may be underperforming (Younus & Qureshi, 2021). By delving deeper into this data and analysing it with increasing levels of granularity, businesses can better identify both problems and opportunities, develop innovative solutions, and capitalize on the strengths of their company (Davenport & Harris, 2007). It is essential for businesses to have a comprehensive understanding of their target audience, as well as other crucial aspects of their operation, to ensure they remain competitive in an ever-evolving marketplace (Gao & Ren, 2020).

This project aims to build a database model using the data of a Lebanese company called “Steelouette”, which will serve as the foundation for creating comprehensive dashboards. The database will provide a unified, scalable, and robust platform to store and manage various types of data, including customer information, sales data, and other crucial metrics.

The project's primary objective is to provide key decision-makers with real-time insights and accurate analysis of their business operations through interactive dashboards. The dashboards will enable users to visualize key performance indicators (KPIs) and other metrics that are essential to monitor business performance (Hernández-Orallo et al., 2020). The dashboards will also provide an intuitive interface for exploring data and discovering trends and patterns that can lead to informed business decisions.

The project will leverage advanced data analytics techniques to create insightful dashboards that provide an overall view of the company's performance, enabling the management team to make informed decisions. It will use a mix of structured and unstructured data, including sales data, customer data, and financial data, among others (Wu et al., 2021). Additionally, the project will employ state-of-the-art data modelling techniques, including entity-relationship diagrams and normalization, to create a robust and scalable database model (García-Sánchez et al., 2018).

Overall, the project aims to create an asset for the company by providing a comprehensive database model and a set of interactive dashboards that will help decision-makers to make informed and data-driven decisions, ultimately improving the company's overall performance.

2. LITERATURE REVIEW

This section presents a literature review about 4 different topics that were considered as essential to this project. The first one would be the importance of the quality of data in any data analysis project. The second talks about the role of big data in the decision making of companies. As for the third one, it highlights the improvement that data makes of efficiency and effectiveness. And last, as we all know, the main goal of a profitable company, is to increase revenue, thus, I focused on the role of data specifically in financial management.

2.1. Quality of Data

In today's business world, data can be generated from multiple sources, such as customer interactions, online reviews, social media analytics, market research, and sales data, among others (Gandomi & Haider, 2015). The challenge for businesses is to consolidate this disparate data into a unified and meaningful form that can be used to drive informed decisions (Breslin, 2019). The process of unifying data requires a deep understanding of the sources, the type of data being generated, and the tools available to analyse it effectively.

Once the data has been unified, businesses must then focus on identifying the specific data points that will be most beneficial for their studies. This requires careful consideration of the key performance indicators (KPIs) that are most relevant to their business objectives, such as customer retention, revenue growth, or market share (Kim & Lee, 2012). By identifying the most valuable data points, businesses can develop more accurate predictive models and optimize their decision-making process.

Moreover, the process of unifying and analysing data is not a one-time event but is an ongoing process that must be regularly reviewed and updated (Alshahrani et al., 2020). With the constant evolution of the business landscape, businesses must stay up to date with the latest trends and technologies to ensure they are leveraging the most relevant data to drive their business forward.

Wieder et al. (2012) highlight the importance of data quality for effective business intelligence. The authors note that while business intelligence tools can provide valuable insights and improve decision-making, they are only as effective as the quality of the data they are analysing. The authors emphasize that poor data quality can lead to inaccurate or incomplete analysis, resulting in flawed decision-making and poor performance. They also note that data quality issues can arise from a variety of sources, such as inconsistent data formats, incomplete data, or incorrect data entry.

To address these issues, the authors suggest that businesses should prioritize data quality management practices, such as data cleansing and validation, to ensure that their business intelligence tools are working with accurate and reliable data. They also recommend regular monitoring and auditing of data quality to identify and address any issues that may arise (Wieder et al., 2012).

Overall, Wieder et al. (2012) highlight the critical importance of data quality for effective business intelligence and underscores the need for businesses to prioritize data quality management practices to achieve optimal performance.

2.2. *Role of Big Data in Reporting and Decision Making*

Fernández, A. (2020) explores the role of big data in reporting and decision-making. With the advent of modern technologies, big data has become a valuable resource for organizations looking to improve their reporting and decision-making processes (Gandomi & Haider, 2015). In terms of reporting, big data can help organizations gather and analyse large amounts of data from various sources to create more accurate and comprehensive reports. These reports can provide valuable insights into customer behaviour, market trends, and other key metrics that can help organizations make more informed decisions (Manyika et al., 2011). By using big data analytics tools, organizations can also streamline their reporting processes, reducing the time and resources needed to produce reports.

When it comes to decision-making, big data can be used to identify patterns and trends in large datasets that would be difficult or impossible to identify using traditional methods (Gandomi & Haider, 2015). By analysing this data, organizations can gain valuable insights into customer behaviour, market trends, and other key factors that can influence business decisions. For example, big data analytics can be used to identify new market opportunities, optimize pricing strategies, and improve customer engagement.

Overall, the role of big data in reporting and decision-making is becoming increasingly important as organizations seek to gain a competitive edge in today's rapidly changing business environment (Manyika et al., 2011). By leveraging the power of big data analytics, organizations can improve their reporting processes and make more informed, data-driven decisions.

2.3. Improvement of Efficiency and Effectiveness

Popescu (2012) explores how business intelligence solutions can help organizations improve their efficiency and effectiveness.

According to Popescu (2012), business intelligence solutions can improve efficiency by streamlining data collection and analysis processes. By automating these processes, organizations can reduce the time and resources needed to collect and analyse data, allowing them to make faster and more accurate decisions. Business intelligence solutions can also help organizations identify inefficiencies in their processes and operations, allowing them to optimize their workflows and reduce waste.

In terms of effectiveness, business intelligence solutions can provide organizations with valuable insights into their customers, markets, and competitors (Popescu, 2012). By analysing data from various sources, organizations can gain a better understanding of customer behaviour, market trends, and other key factors that can impact business performance. This information can be used to develop more effective marketing and sales strategies, improve customer engagement, and identify new business opportunities.

Popescu (2012) also notes that business intelligence solutions can help organizations improve their decision-making processes. By providing decision-makers with real-time access to data and insights, organizations can make more informed decisions that are based on accurate and up-to-date information. Business intelligence solutions can also provide decision-makers with tools and visualizations that make it easier to interpret complex data and identify trends and patterns.

Overall, the use of business intelligence solutions can lead to general improvements in efficiency and effectiveness for organizations (Popescu, 2012). By leveraging the power of data and analytics, organizations can optimize their processes, improve their decision-making capabilities, and gain a competitive edge in today's rapidly changing business environment.

2.4. Role in Financial Management

Bray (2011) explores how business intelligence dashboards can help organizations improve their financial management processes.

According to Bray (2011), business intelligence dashboards can provide financial managers with a real-time view of their organization's financial performance. This allows them to quickly identify potential issues or areas of concern and take action to address them.

In addition to providing real-time insights into financial performance, business intelligence dashboards can also improve collaboration and communication within an organization. By providing a centralized view of financial data, Bray (2011) explains that dashboards can facilitate collaboration between different departments and teams. This can lead to more effective communication and decision-making, as all stakeholders have access to the same information.

Another benefit of business intelligence dashboards is that they can be customized to meet the specific needs of an organization. Financial managers can choose which KPIs and metrics to display on the dashboard, based on their unique needs and goals. This flexibility allows organizations to tailor their financial management processes to their specific requirements, improving efficiency and effectiveness (Bray, 2011).

Overall, according to Bray (2011), the use of business intelligence dashboards can play a crucial role in financial management. By providing real-time insights, facilitating collaboration and communication, and offering flexibility and customization, dashboards can help organizations improve their financial performance and achieve their goals.

3. METHODOLOGY

In this section, I am going to describe the process that was followed throughout this project, starting by the exploration of the data, explaining the choice of the tools, giving the sense behind the data model, highlighting the main steps in the ETL process, and finally showing the look and organization of the dashboards produced during the project.

3.1. Datasets

The company "Steelouette" currently records their sales and other data manually using a cloud-based spreadsheet software called "Airtable". In order to process the data, files had to be exported manually and be labelled based on their content. Each file was renamed following a specific format for easy integration. The first six digits represent the year and month, while the remaining characters signify the category of data within the file. There were eight categories: non-customized, customized, events, special stores, wine, crib, keychains, and bowties.

An evaluation has been conducted on the data to identify the main data points that are needed, their value ranges, and their usefulness for studying the data. No data cleaning has been performed prior to this evaluation; all changes will be made in later steps.

To note, the data the author worked with dates from February 2021 till December 2022.

3.2. Tools

To perform the ETL process, two tools were chosen, Python and PostgreSQL. Those 2 tools have been mainly picked for their ease of use to manipulate and clean the data. The csv files were integrated into python, and through a connection to the PostgreSQL were integrated in the tables.

Python is a popular and powerful programming language for data integration and analysis due to its large ecosystem of libraries and tools, ease of use, versatility, and great community support. Python has libraries such as NumPy and Pandas, and that can help to easily manipulate data. It is easy to learn and use. Python is versatile and can be used for a wide range of tasks, beyond just data integration.

As for PostgreSQL, it is a powerful and highly scalable open-source relational database management system that can handle large amounts of data and high transaction volumes. Its rich set of built-in data types and advanced data manipulation features make it versatile and customizable for specific needs. Additionally, it provides strong data security features and has a large and active community of developers who provide support for users. Its extensibility allows for customization through a large number of third-party extensions and add-ons, making it a popular choice for data manipulation projects.

For building the dashboards, Microsoft Power BI was used. It is an easy-to-use and flexible business intelligence tool that allows users to create interactive dashboards and reports. It integrates well with other Microsoft products and data sources, provides robust data modelling and analysis capabilities. As a cloud-based platform, it offers scalability and automatic updates and maintenance. Its large and active user community and Microsoft's extensive documentation and support resources make it a popular choice for businesses seeking an intuitive and powerful tool for data visualization and analysis.

That way, the whole flow that usually data passes through was covered.

3.3. Data Model

This section will cover the data model that was built for the company, containing two main areas: revenue and customer satisfaction.

To build the model, it was decided to use a star schema, a database schema used in data warehousing that organizes data into a central fact table surrounded by dimension tables. This schema simplifies queries, improves performance, and provides scalability. The design is simple and easy to maintain, making it efficient for large data sets.

The model includes two fact tables: Revenue and Comments. The Revenue table will contain information about each product sold, while the Comments table will hold the number of comments per order per designer, allowing them to evaluate customer satisfaction and track employee performance. The main dimension tables will be Date, Products, Types, Stores, Gender, Customers, and Designers, providing the flexibility to analyse revenue numbers from different perspectives and make better marketing and workflow decisions to maximize performance and profit.

As for the comments table, it will hold the number of comments per order per designer, so they could evaluate the satisfaction of the customers and eventually track employees that are causing delays on several occasions.

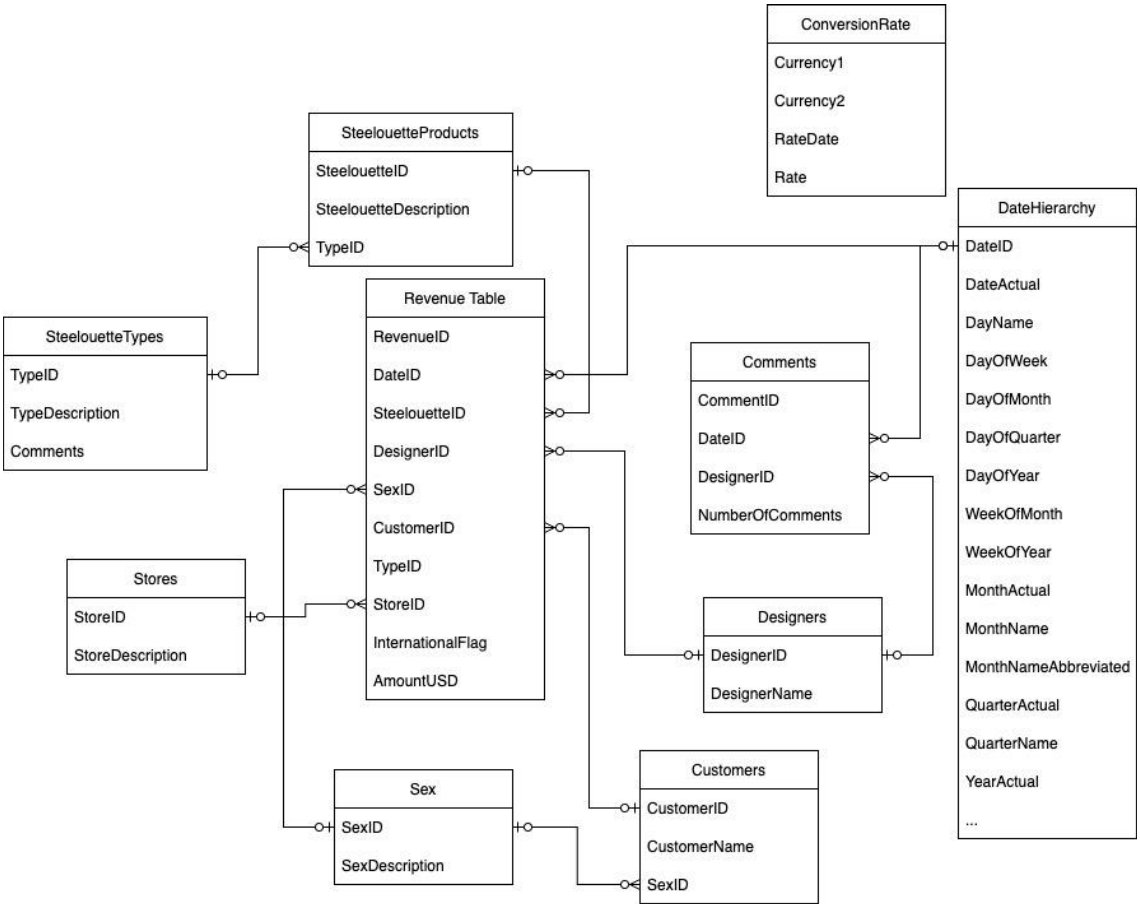

Figure 1: Database Model Built for “Steelouette”

The tables follow normalization rules, and each of them has a logical PK as well as a physical one created by a sequence that increments by one for each record inserted. Additionally, foreign keys (FK) have been created to ensure that each record being inserted has the correct linked details.

The data files contained additional information, but the quality was not sufficient for inclusion. This topic will be further discussed in the results section and the expansion part.

3.4. ETL Process

The ETL process is divided into 3 main parts. The chronology is essential for 2 important reasons: the first one, it ensures that the tables are being filled in an order that would fill the dimensions tables before the fact ones so that there is no violation of the rules of the FK. As for the second one, it divides the process in different steps to help identify faster, possible errors and integration issues.

The first part is an SQL script, that would create all the tables that are part of the model, as well as, staging tables that will include the first version of the raw data before being transformed and elaborated. In that chunk, several DDL queries are performed to create PK, FK, and indexes.

```sql
Create Table Steelouette_Revenue
(
    RevenueId           SERIAL Primary Key,
    DateId              int,
    SteelouetteId       int,
    SexId               int,
    DesignerId          int,
    CustomerId          int,
    TypeId              int,
    StoreId             int,
    InternationalFlag   int,
    AmountUSD           float
);

Create Table Comments
(
    CommentId           SERIAL Primary Key,
    DateId              int,
    DesignerId          int,
    NumberOfComments    float
);

ALTER TABLE Steelouette_Revenue
    ADD CONSTRAINT fk_revenue_type FOREIGN KEY (TypeId) REFERENCES Steelouette_type (TypeId);

ALTER TABLE Steelouette_Revenue
    ADD CONSTRAINT fk_revenue_store FOREIGN KEY (StoreId) REFERENCES Steelouette_Stores (StoreId);

```

Figure 2: DDL commands to create table, PK, and FK.

The second step is a Python script that has as objective extracting the data from the csv files into data frames using pandas. First, the rate conversion data is integrated. This data has been exported from a website called lirarate.org[1] into a csv, and then inserted into the rate table on the database using a connection to Postgres.

```
In [2]: df_rates = pd.read_csv("/Users/gerardnajem/Desktop/ISEG/MFW/usd-to-lbp-market-rate.csv")
        df_rates.rename(columns = {'USD to LBP':'Rate'}, inplace = True)
        df_rates.rename(columns = {'DateTime':'RateDate'}, inplace = True)
        df_rates.insert(0, column='Currency2', value='LBP')
        df_rates.insert(0, column='Currency1', value='USD')

In [3]: conn = psycopg2.connect(
            database="Steelouette", user='postgres', password='011095', host='localhost', port='5432'
        )

In [4]: def execute_values(conn, df, table):

            tuples = [tuple(x) for x in df.to_numpy()]

            cols = ','.join(list(df.columns))
            # SQL query to execute
            query = "INSERT INTO %s(%s) VALUES %%s" % (table, cols)
            cursor = conn.cursor()
            try:
                extras.execute_values(cursor, query, tuples)
                conn.commit()
            except (Exception, psycopg2.DatabaseError) as error:
                print("Error: %s" % error)
                conn.rollback()
                cursor.close()
                return 1
            print("the dataframe is inserted")
            cursor.close()

        execute_values(conn, df_rates, 'Conversion_rate')
```

Figure 3: Python script to integrate conversion rate data.

The same process has been applied to the csv files containing the data, however it was divided into 8 different part, since each type of file need to be linked to its specific information and labelled to simplify the process later on. The needed columns were selected from the data frames, others were merged together, and then integrated each in its specific staging table.

[1] Here it is important to indicate that there is no official data regarding the USD rate in the Lebanese market, since the state still did not officially claim officially, thus, the only rate is the black market one. However, it is the widely used rate for goods and services.

```python
for i in csv_files:
    if re.search('R',i):
        if R.empty:
            string = "/Users/gerardnajem/Desktop/ISEG/MFW/Data/" + i
            R=pd.read_csv(string)
            R.insert(0, column='Type', value='Non-Customized')
            R.insert(0, column='Week', value='1')
            R.insert(0, column='Month', value= i[4:6])
            R.insert(0, column='Year', value= i[0:4])
        else:
            string = "/Users/gerardnajem/Desktop/ISEG/MFW/Data/" + i
            R1=pd.read_csv(string)
            R1.insert(0, column='Type', value='Non-Customized')
            R1.insert(0, column='Week', value='1')
            R1.insert(0, column='Month', value= i[4:6])
            R1.insert(0, column='Year', value= i[0:4])
            R=pd.concat([R1,R])
    elif re.search('Bowties',i):
        if KeyChains.empty:
            string = "/Users/gerardnajem/Desktop/ISEG/MFW/Data/" + i
            Bowties=pd.read_csv(string)
            Bowties.insert(0, column='Type', value='Bowties')
            Bowties.insert(0, column='Week', value='1')
            Bowties.insert(0, column='Month', value= i[4:6])
            Bowties.insert(0, column='Year', value= i[0:4])
        else:
            string = "/Users/gerardnajem/Desktop/ISEG/MFW/Data/" + i
            Bowties1=pd.read_csv(string)
            Bowties1.insert(0, column='Type', value='Bowties')
            Bowties1.insert(0, column='Week', value='1')
            Bowties1.insert(0, column='Month', value= i[4:6])
            Bowties1.insert(0, column='Year', value= i[0:4])
            Bowties=pd.concat([Bowties1,Bowties])
```

Figure 4: Python script to integrate the revenue and comments data.

The last and third part is the actual data transformation and manipulation. It starts by filling the dimension tables; some of them contain static data such as the sex, the date hierarchy, the stores, and the types. The others, such as customers, products are filled by checking the information that was collected from the csv files dynamically. Afterwards, several transformations are performed on the data to ensure consistency and identify outliers and data points that do not present any additional value to the analysis[2].

```sql
INSERT INTO Sex (SexDescription) values ('Female');
INSERT INTO Sex (SexDescription) values ('Male');
INSERT INTO Sex (SexDescription) values ('Unknown');

INSERT INTO Steelouette_type (TypeDescription, Comments) values ('Customized','');
INSERT INTO Steelouette_type (TypeDescription, Comments) values ('Bowtie','');
INSERT INTO Steelouette_type (TypeDescription, Comments) values ('KeyChain','');
INSERT INTO Steelouette_type (TypeDescription, Comments) values ('Wine','');
INSERT INTO Steelouette_type (TypeDescription, Comments) values ('Crib','');
INSERT INTO Steelouette_type (TypeDescription, Comments) values ('Non-Customized','');
INSERT INTO Steelouette_type (TypeDescription, Comments) values ('Unknown','');
```

Figure 5: Filling some dimension tables statically.

[2] Here are some snippets of the code, otherwise, the rest of the code can be sent upon request.

```sql
INSERT INTO Steelouette_products (SteelouetteDescription, TypeId)
SELECT distinct INITCAP(TRIM(unnest(string_to_array(r.steelouettename, ',')))),
        (select typeid from steelouette_Type where typedescription = 'Bowtie')
  FROM Bowties_TEMP r
  where r.steelouettename <> 'Unknown'
UNION ALL
SELECT distinct INITCAP(TRIM(unnest(string_to_array(r.steelouettename, ',')))),
        (select typeid from steelouette_Type where typedescription = 'KeyChain')
  FROM Keychains_TEMP r
  where r.steelouettename <> 'Unknown'
UNION ALL
SELECT 'Wine',
        (select typeid from steelouette_Type where typedescription = 'Wine')
UNION ALL
SELECT 'Crib',
        (select typeid from steelouette_Type where typedescription = 'Crib')
```

Figure 6: Filling some dimension tables dynamically.

This step is the most important one, since it ensures having clean, correct, and understandable data, that would eventually facilitates building the dashboard and extract the insights and conclusions from the project.

```sql
CREATE TABLE R_TEMP AS
SELECT CAST(R.Year       AS INT) AS Year,
       CAST(R.Month      AS INT) AS Month,
       CAST(R.Week       AS INT) AS Week,
       CAST(COALESCE(NULLIF(SUBSTRING(R.Date from 1 for
                (CASE WHEN POSITION(',' in R.Date) = 0 THEN length(R.Date)
                      ELSE POSITION(',' in R.Date) - 1
                      END
                )
               ), ''),'1/1/1970') AS DATE)  AS Date,
       CAST(INITCAP(TRIM(R.ClientName)) AS VARCHAR(100))  AS ClientName,
       CAST(COALESCE(NULLIF(R.SteelouetteName,''),'Unknown') AS VARCHAR(100)) AS SteelouetteName,
       CAST(CASE WHEN TRIM(R.Boy) = 'checked' THEN 'Male'
            WHEN TRIM(R.Girl) = 'checked' THEN 'Female'
            ELSE 'Unknown'
            END                       AS VARCHAR(10)) AS Sex,
       CAST(COALESCE(NULLIF(R.Number, ''),'0') AS FLOAT) AS Number,
       CAST(CASE WHEN NULLIF(R.SteelouetteName,'') like '%bowtie%' THEN 'Bowties'
                 WHEN NULLIF(R.SteelouetteName,'') like '%keychain%' THEN 'KeyChain'
            ELSE R.Type END AS VARCHAR(100))        AS Type,
       CAST(CASE WHEN position('DHL' in R.DelOpt) <> '0' THEN 'International' ELSE 'Local' END AS Varchar(20)) DelOpt,
       CAST(NULLIF(REGEXP_REPLACE(REGEXP_REPLACE(NULLIF(R.ValueUSD,''), '[^.0-9]+', '', 'g'), E'\\r?\\n', '', 'g'),'') AS FLOAT)
       CAST(NULLIF(REGEXP_REPLACE(REGEXP_REPLACE(NULLIF(R.Value,''), '[^.0-9]+', '', 'g'), E'\\r?\\n', '', 'g'),'') AS FLOAT) AS
  FROM StagingR R;
```

Figure 7: Example of data manipulation and transformation.

3.5. Dashboards

To wrap up the project, dynamic dashboards were built to contain the main results of the study and it has been divided into 3 main parts.

The first one is an overview of the business, with the main numbers, and trends. The second page contains the same numbers but sliced into categories, to highlight the success and failure of main products, as well as underline the trend of some other products in specific seasons. As for the last and third page, the goal was to show the importance of the quality of data, by emphasizing on the missing values that might impact the conclusions that we gather.

Most of the calculations have been made before connecting Power BI to the database, however, others had to be made using DAX while developing the different tiles and pages.

The dashboard[3] has been built in a dynamic way, so the user could select a period of time, could slice the numbers by clicking on a specific time, or even drag one into the date hierarchy, going from yearly figures, to quarterly, monthly, or even daily ones.

[3] Refer to the appendices section to check the dashboards..

4. RESULTS

In this section, the list of the main results that have been collected will be presented. These results were sent to the company to provide understandings on the several products, types, and stores that they collaborate with. On top of that, the author offered several tips to improve the data, and ultimately the analysis of the production.

Starting by the overview page, aside from seeing the revenue numbers and the number of items sold, it's possible to see the percentage of this share coming from international sales. This number has an important value, since the economic situation in Lebanon is considered as being unstable, any revenue coming from abroad, makes the business more reliable and steadier. This is clear since 3% of the items sold internationally are bringing almost 12% of the revenue.

The numbers also show that females are the main customers for the business since they form 48% of the total sales, but it's important to note that around 40% of sold items don't specify the gender, which could affect the numbers shown.

It's also obvious, that the peak of sales comes at the end of the year before the Christmas period, as the months of November and December provides the highest numbers in term of sales and revenue. A slight increase can be seen as well during summer where sales are impacted by expats coming in and out of Lebanon and buying gifts to their friends and family.

One last table was added to this page, and it was requested by the company itself. It presents the average of comments received by customers for each customized item produced. It is sliced by designers that worked there, to track employees that might be causing a dissatisfaction for customers.

Now, looking at the detailed view, several outcomes can be gathered. First, around 82% of the products sold, are customized ones, and 92% of the sales are made either through social media, or in-store. Excluding those two numbers, an overview of the top 5 products sold and the top 5 stores are presented in a tree map to emphasise on the main income streams.

Another graph has been added to check the evolution of specific categories of products or new collections that have been released. Let's take as an example the

"Bowtie" collected which was launched in Spring 2022, and we can see its benefits since it peaked in summer during the wedding season. Another one is the "Crib" collection that can be only seen in December since it is produced for Christmas. Non-customized products are being produced all along the year, however, it can slightly be seen that there is an increase in March, before Mother's Day, and at the end of the year.

As cited before, the main purpose of the last page was to show the gaps in the data. 95% of the revenue is not being linked to any specific product, and 95% of the items sold have not been assigned any product. In addition, it's important to mention it again, 41% of sales is assigned to customers with no gender, however it composes only 7% of the revenue. Moreover, 13% of sales and 12% of revenue come from products with no category (which is due to sales in other stores, where data is not being collected in the right way).

Other miscellaneous information was noticed, such as 20% of the items designed have an unknown designer, and 5% of the products sold don't have a clear customer.

All those gaps might not prove to be impactful on the business, but fixing it, would benefit it and raise the quality of the analysis that is being made.

5. EXPANSION AND FUTURE WORK

In this section, the author is going to cover the future work that could be completed to improve the effort made in this project. This is going to be proposed and discussed with the company in the following months.

One important point, as mentioned before, the data entry flow should be optimized and improved to keep the consistency in the information added as well as raise the quality of the data that is being looked at. As discussed in the literature review, there was an emphasis on the importance of having data quality, since it can affect the accuracy and the value of the insights that we are getting. Thus, the first step would be to create a unified, central flow and documentation for data entry.

After improving the above, more information could be collected that could help study more data points and therefore build more personalized KPIs and dashboards that would benefit either the marketing side, or the performance of work in the company.

Since the company is based on selling products to individuals mainly, first step would be to focus on information related to these customers. Maybe start by building profiles for these customers that goes beyond the name and the sex, that way they would be able to use this analysis to target specific range of customers in their marketing campaigns.

Another thing that could be added is having a whole dictionary for the products available, by clearly dividing them into categories and subcategories, and this would lead to have a clearer view of what are the most profitable items and collections in the business.

Last, this work could be expanded to include the internal data as well, such as employees, the workflow that each product follows, the duration from production till delivery and all the steps that it includes. This would help to eventually catch the bottle necks that might be affecting the performance and therefore augment the sales and revenue.

6. CONCLUSION

In conclusion, the results gathered from the analysis of the company's data provide valuable insights into the products, types, and stores the business collaborates with. The percentage of revenue generated from international sales highlights the importance of diversifying revenue streams and the data also shows that females are the main customers, and the peak of sales comes during the end of the year and summer months. The detailed view reveals that customized products make up the majority of sales, and social media and in-store purchases are the main sales channels.

Many of those conclusions are visible in the dashboards, and that’s the importance of business intelligence in financial management. As discussed in the literature review section, these analysis have an impact on both the revenue streams and the effectiveness and efficiency of the workflow in the company, and all these points, can be tracked in the dashboards right now, or when more data points will be added.

Moreover, the gaps in the data, such as unidentified products, designers, and customers, may not have an immediate impact, but addressing them will improve the quality of the analysis and ultimately benefit the business. Thus, this step is as essential as adding new information to the model, since the quality of data help achieve optimal performance as discussed in the second section of this report.

Overall, this analysis provides a roadmap for improving the business and increasing revenue in the future by building business intelligence tools that could help track on a real time basis all the information that is needed in its smallest detail.

REFERENCES

Alshahrani M, Gondal I, Iqbal R, Zeadally S. Continuous Data Analysis: A Practical Approach for Precision Medicine. IEEE Access. 2020.

Bray, T. (2011). The Role of Business Intelligence Dashboards in Financial Management. Credit Control , 32 (5/6), 29–33.

Breslin, J. G. (2019). Developing an organizational data strategy. Journal of Business and Management, 24(1), 35-42.

Cukier, K. (2013). Data, data everywhere: A special report on managing information. The Economist, 25.

Davenport, T. H., & Harris, J. G. (2007). Competing on analytics: The new science of winning. Harvard Business Press.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. National Science Review, 1(2), 293-314.

Fernández, A. (2020). El papel del Big Data en el reporting y la toma de decisiones. Contabilidad y Direccion, 31, 21–36.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137-144. doi: 10.1016/j.ijinfomgt.2014.10.007

Gao, J., & Ren, G. (2020). Social media data analytics: An overview. IEEE Transactions on Computational Social Systems, 7(6), 1621-1633.

García-Sánchez, F., Pérez-González, D., & Pavón-Mariño, P. (2018). A survey of data modeling techniques for big data. Information Systems, 73, 1-19.

Hernández-Orallo, J., Ramírez-Quintana, M. J., & Flach, P. A. (2020). Explainable AI: Interpreting, explaining and visualizing deep learning. Springer.

Kim, S. H., & Lee, J. N. (2012). The impact of organizational context and information technology on employee knowledge-sharing capabilities. Public Administration Review, 72(4), 527-536.

Kudyba, S., & Hoptroff, R. (2017). Data-driven: Creating a data culture. Apress.

Li, X., Huang, J., Wang, L., & Lu, X. (2018). Big data analytics in supply chain management: A comprehensive overview and future directions. Computers & Industrial Engineering, 116, 299-316.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, 1-156.

Noori, H., Ebrahimi, M., & Moghimi, S. M. (2020). The role of big data analytics in improving supply chain management. Journal of Enterprise Information Management.

Popescu, S. (2012). Business Intelligence Solutions - a Way of General Improvement of Efficiency and Effectiveness. Review of International Comparative Management / Revista de Management Comparat International, 13(1), 88–95.

Wieder, B., Ossimitz, M.-L., & Chamoni, P. (2012). The Impact of Business Intelligence Tools on Performance: A User Satisfaction Paradox? International Journal of Economic Sciences & Applied Research, 5(3), 7–32.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2021). Data mining with big data. IEEE Transactions on Knowledge and Data Engineering, 33(5), 1855-1878.

Younus, A., & Qureshi, M. A. (2021). Data Analytics for Business Decision Making: A Comparative Study of Regression Techniques. In Proceedings of the 4th International Conference on Future Networks and Distributed Systems (pp. 229-241).

APPENDICES

Figure 8: Overview Tab in dashboard

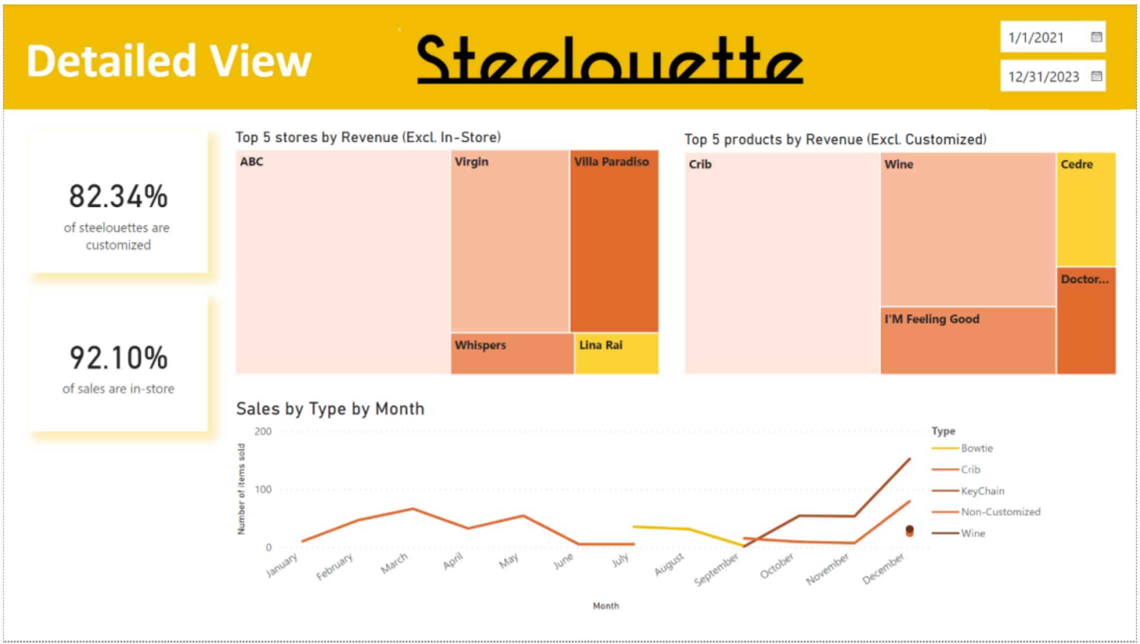

Figure 9: Detailed View in dashboard

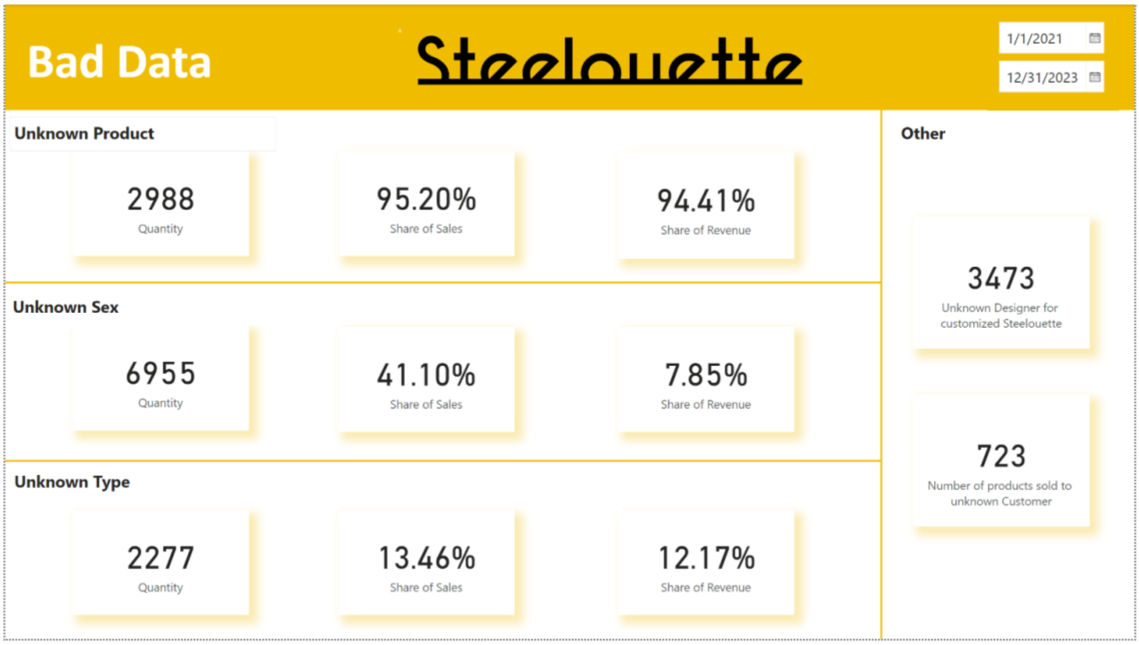

Figure 10: Bad Data View in dashboard