ADVENTURE WORKS: COMPANY ANALYSIS

Epale! In this text you can see my step by step to generate a complete interactive report with 7 panels and an analysis of the last years of the company, this is an end-to-end project. You can see the report in this link.

About this project

Adventure Works is the example dataset in Microsoft. I used that dataset for make a real company report. Using **SQL Server in SSMS**, making some different data models, a querys with JOINs for Views. Using **Power BI**, load the data through Views in connection with SQL Server, make a calendar table in **Power Query** and others transformations about the data, make some measures in **DAX**, and a start model. Then I perform an analysis of the data obtained in the report. Using **Figma** I make the backgrounds in the report.

About the data

This data is free to use, you can download that in this link and read more about that if you want. I use the AdventureWorks2019.bak

Steps

I divide that document in 2 parts, the first part is why I make the report, and the second part is my analysis about the report.

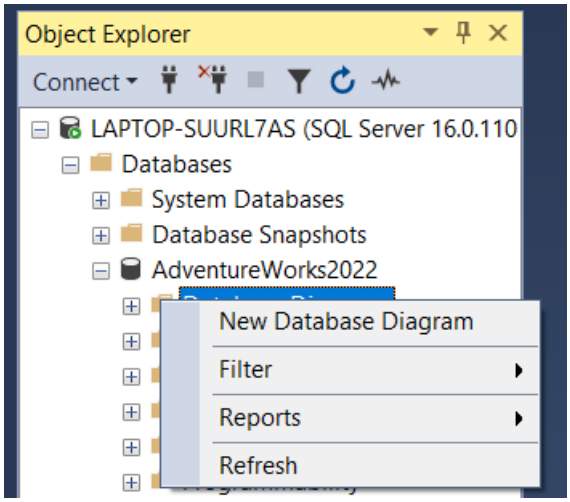
MAKE THE REPORT

AdventureWorks is a fictitious company, where we can obtain data as if it were a real company, purchasing data, sales, deposit information, productions, customers, suppliers, among others.

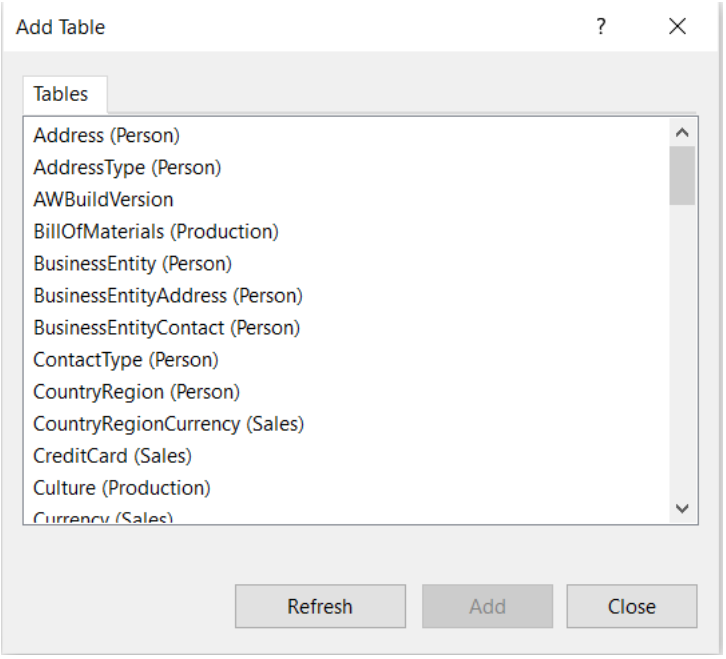
The first task we would have been to ask ourselves questions about the data, what we are going to obtain with it and what we want to know about it, but to do this we must know what is inside the database, so we must perform an EDA in SSMS (SQL Server).

EDA & ETL

Once we are connected to **SQL Server**, we go to make the news database diagrams:



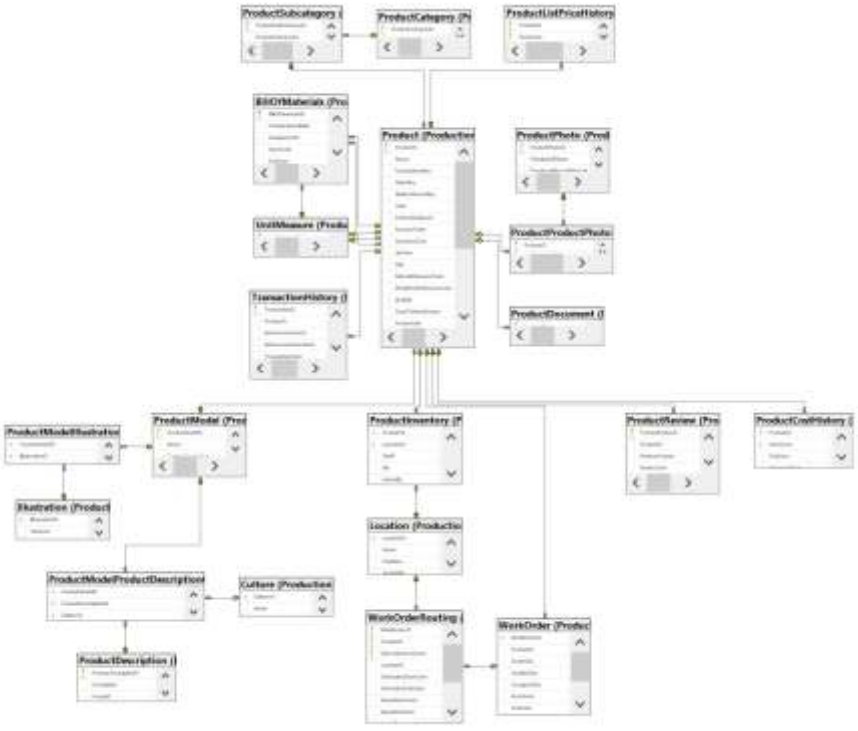
We are lucky that this database is all documented and well made, so when creating database diagrams, it gives us the following window:



Here you can see each of the tables and in parentheses the scheme to which it corresponds, such as: Person, Production, Sales and others. Thanks to this we will create the different diagrams: Person, Production, Purchases, HR and Sales.

NOTE: If you did not have this you would have to test between different tables to see which fields you can make a connection between.

Then you must see the different diagrams, because even if you have created them, there are still many tables and even more fields. Here is an example of the Production diagram:



Throughout these diagrams you must review which fields you think will be used for the subsequent analysis, this is an extensive selection work, carrying out queries to obtain said information, to be able to define your Facts table, your Dimensions tables and the fields you want of them.

After defining that, you must perform the final query. For the Production table it will look like this:

```
SELECT
    wk.WorkOrderID,
    wk.StartDate,
    wk.EndDate,
    wk.DueDate,
    l.LocationID,
    l.Name AS Location,
    wk.ProductID,
    p.Name AS Product,
    pm.ProductModelID,
    pm.Name AS ProductModel,
    wk.OrderQty,
    wk.StockedQty,
    p.StandardCost,
    p.ListPrice AS UnitaryPrice
FROM
    Production.WorkOrder AS wk
    INNER JOIN
        Production.Product AS p ON wk.ProductID = p.ProductID
    INNER JOIN
        Production.WorkOrderRouting AS wkr ON wk.WorkOrderID = wkr.WorkOrderID
    INNER JOIN
        Production.Location AS l ON wkr.LocationID = l.LocationID
    INNER JOIN
        Production.ProductModel AS pm ON p.ProductModelID = pm.ProductModelID
```

You can test the query to make sure it doesn't give you an error:

Now to generate the view you must place at the beginning:

```
CREATE VIEW Production_PowerBI AS
```

And then all the query described above. With this you will have the first View saved as Production_PowerBI

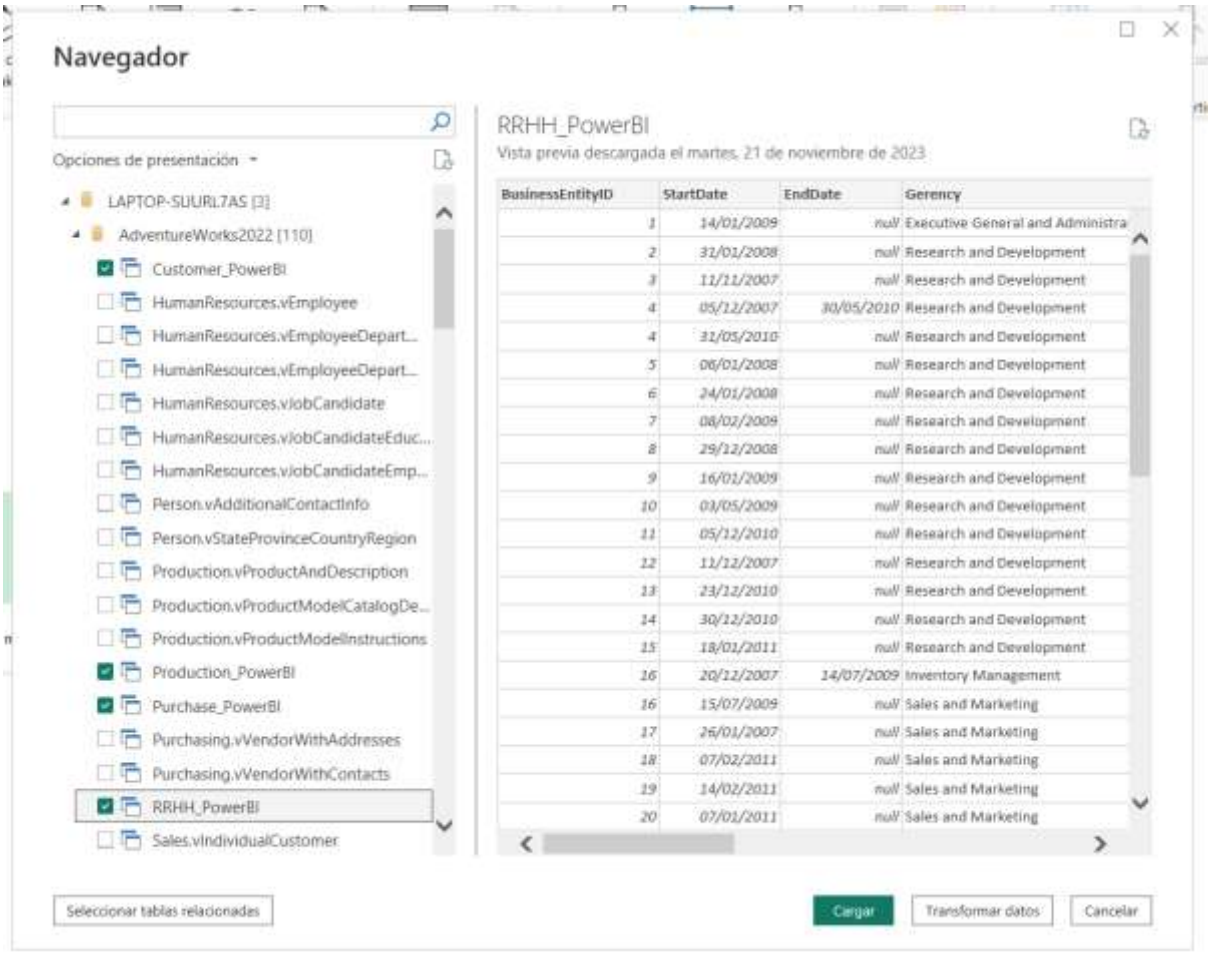
After carrying out this process with the Sales, HR, Purchases, Stock (Depot) and Customers views, we go to **Power BI Desktop.** Once the program is open we must obtain the data from **SQL Server**:

NOTE: Sorry because the language is in Spanish, it is my native language and I feel more comfortable working in it.

This open that window:

You must write the name of the server in SSMS, which in my case is "LAPTOP-SUURL7AS".

Now select the corresponding Views:



You can load the data at once, but it is better to transform it (Transformar datos), data for define the data, like StartDate is date data type, and Gerency is text data type, and others the same. This is to maintain data integrity, is very important in the reports, because it helps to avoid making mistakes in its operation.

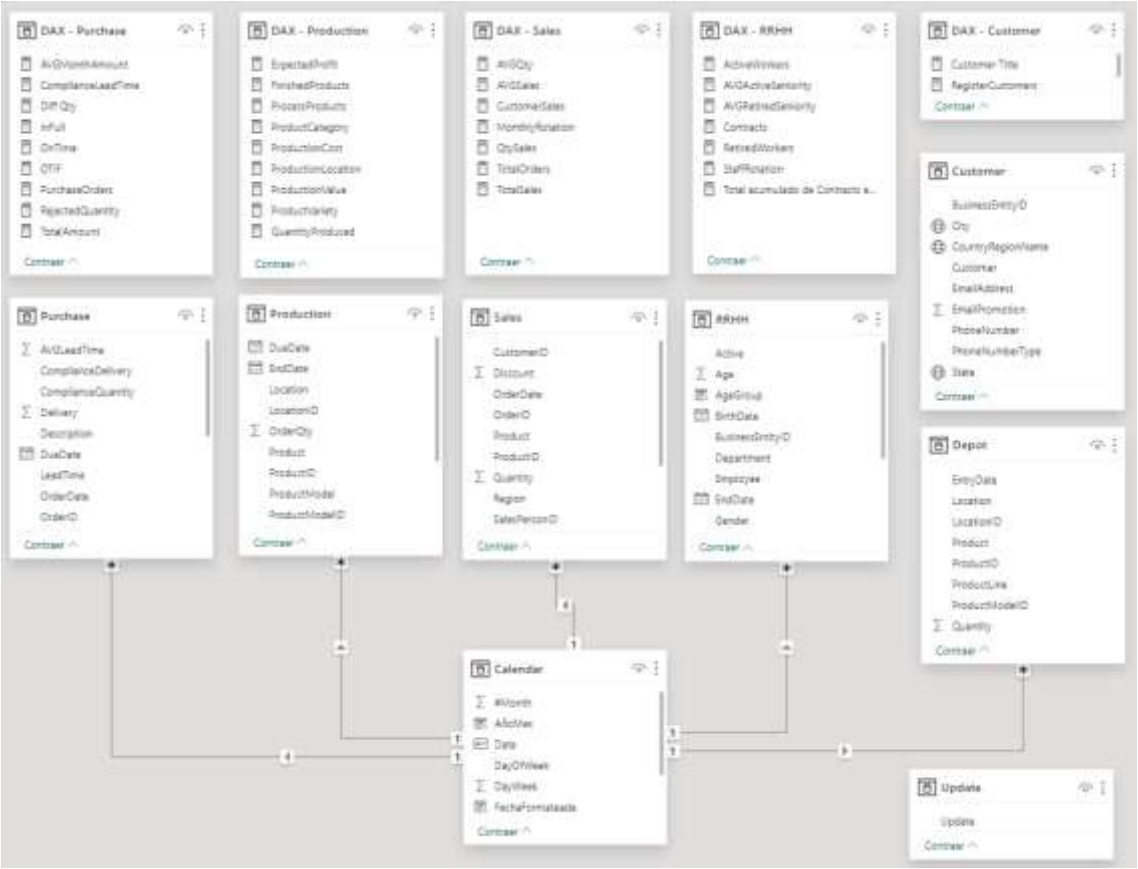
Here we add some columns, do some subtraction of dates in M, other conditionals, a bit of data transformation. This process is a bit tedious so I won't show it, but here is the script I used for the Calendar table:

```
let

  //Get Max and Min date from oders table

  //MaxDate=(Date.From(DateTime.LocalNow())),

  //MaxDate=List.Max(RRHH[StartDate]),

  MaxDate=List.Max(Depot[EntryDate]),

  MinDate=List.Min(RRHH[StartDate]),

  //MinDate=Date.FromText("2021/01/01"),

  NoDays= Number.From(MaxDate-MinDate)+1,

  Origen = List.Dates(MinDate, NoDays, #duration(1, 0, 0, 0)),

  #"Convertida en tabla" = Table.FromList(Origen, Splitter.SplitByNothing(), null, null, ExtraValues.Error),

  #"Columnas con nombre cambiado" = Table.RenameColumns(#"Convertida en tabla",{{"Column1", "Date"}}),

  #"Tipo cambiado" = Table.TransformColumnTypes(#"Columnas con nombre cambiado",{{"Date", type date}}),

  #"Columna duplicada" = Table.DuplicateColumn(#"Tipo cambiado", "Date", "Date - Copia"),

  #"Columnas con nombre cambiado1" = Table.RenameColumns(#"Columna duplicada",{{"Date - Copia",
"DayOfWeek"}}),

  #"Nombre del día extraído" = Table.TransformColumns(#"Columnas con nombre cambiado1", {{"DayOfWeek", each
Date.DayOfWeekName(_), type text}}),

  #"Día de la semana insertado" = Table.AddColumn(#"Nombre del día extraído", "DayWeek", each
Date.DayOfWeek([Date]), Int64.Type),

  #"Nombre del mes insertado" = Table.AddColumn(#"Día de la semana insertado", "Month", each
Date.MonthName([Date]), type text),

  #"Año insertado" = Table.AddColumn(#"Nombre del mes insertado", "Year", each Date.Year([Date]), Int64.Type),

  #"Columnas combinadas" = Table.CombineColumns(Table.TransformColumnTypes(#"Año insertado", {{"Year", type
text}}, "es-419"),{"Month", "Year"},Combiner.CombineTextByDelimiter(" ", QuoteStyle.None),"MonthYear"),

    #"Nombre del mes insertado1" = Table.AddColumn(#"Columnas combinadas", "Nombre del mes", each
Date.MonthName([Date]), type text),

    #"Año insertado1" = Table.AddColumn(#"Nombre del mes insertado1", "Año", each Date.Year([Date]),
Int64.Type),

    #"Columnas con nombre cambiado2" = Table.RenameColumns(#"Año insertado1",{{"Nombre del mes", "Month"},
{"Año", "Year"}}),

    #"Mes insertado" = Table.AddColumn(#"Columnas con nombre cambiado2", " #Month", each Date.Month([Date]),
Int64.Type)

in

    #"Mes insertado"
```

And Update table:

```
let

    Origen = DateTime.LocalNow(),

    #"Convertida en tabla" = #table(1, {{Origen}}),

    #"Columnas con nombre cambiado" = Table.RenameColumns(#"Convertida en tabla",{{"Column1", "Update"}})

in

    #"Columnas con nombre cambiado"
```

Now with all the data loaded into Power BI and transformed into Power Query, we can move on to data modeling within Power BI:



This is the final model, it's a Star model with the respective DAX table. I made a table of measurements for each of the areas that were necessary (all except Depot). In this document I don't look all syntax in mesurements because are a lot (+40), but I use different functions:

```
SELECTEDVALUE()
DISTINCTCOUNT()
DIVIDE()
SUM()
CALCULATE()
FILTER()
COUNTROWS()
ALLSELECTED()
```

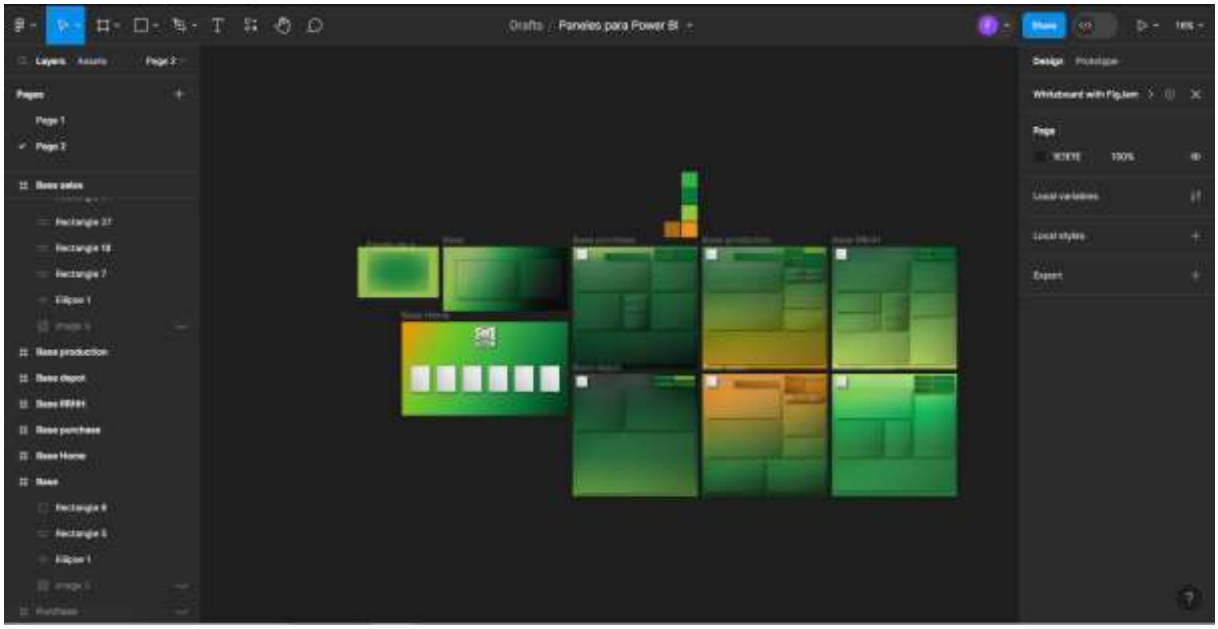
And others.

With that I made the reports, seven interactive dashboards, with multiple graphs.

Backgrounds images

I made the backrounds in Figma, I use the colors in the logo:

Those are my drafts:



One example, the color palette, home, and all areas.

The logo is from windows page, but I was to edit, because it's in JPG on the page. The edition was in Photoshop.

REPORT (DASHBOARDS)

The report is divided into six different areas, where graphs and tables are shown for each one in particular, so these will vary depending on each one and the objectives that each area has. These objectives will be explained in the analysis, only images of the reports will be presented here.

Purchase

Production

Depot


Flow of product entries to the depot
02/12/2023 20:18:26
Quantity per Year
Number of products per Month/Year
Entries of finished products to the depot

Sales


Sales Analytical Summary
02/12/2023 20:18:26
Sales per Date
Product Detail
Region
Sales per Seller
Orders & Sales per Customer
Orders & Sales per Region

HR


Analysis of Human Talent management
02/12/2023 20:18:26
Contracts by Gerency
Contracts per Date
Contracts classified by Gender
% Contracts by Age
AVG Active Seniority by Gerency
AVG Active Seniority by Department
Contract & Seniority per Job

Customer


Summary of the Portfolio of registered Clients (18484) in the world
02/12/2023 20:18:26
Customers by Country
Register Customers per State
Customers by Cities
Number of promotional emails per Customer (%)
Customer information

ANALISIS ABOUT THE REPORT

Here I define all áreas of the report, with the objetives and, why I make in this way.

In this report, a descriptive analysis of the different areas was carried out, where each one has objectives of interest.

Purchase

This area is dedicated to the purchase of raw materials and supplies that will be necessary for the sale. Goals:

1) The main objective was to know how much was spent and which companies bought the most from.
2) And we wanted to know which companies were failing us with deliveries or with the quantities in them (OTIF).

To make these objectives a reality we used the scheme within Purchase, where from all the orders (4011) we were able to calculate the total amount of purchases ($115.84M) and a monthly average ($1.17M). Thus, creating the following graphs:

- Total Amount per Date: Here you can see the money spent on resources over time.
- Total Amount per Supplier: The money spent on each supplier and which suppliers it is spent on the most.
- IF&OT per Supplier: Which suppliers have problems with IF&OT.
- IF&OT per Product: And with which particular products the IF&OT problems occur.

Then we have the **OTIF**, this helps measure the delivery capacity and quantity. We can see that the delivery is quite good, but in quantity it may have slight problems. In particular with the Sport-100 Helmet product in its three colors.

NOTE: The OTIF measure is normally used to see sales, if our deliveries arrived on time and with the correct quantity, but since I was unable to do so due to the information, I implemented it here to put it into practice.

Here you can also see the three largest providers and even the rest in order of total amounts. Finally, there is a table where you can view product by product, differentiating by supplier, where the product will be repeated and the supplier will change.

Production

For this production report it was important to know about:

1) The processed products and the finished products, the total quantity of each at the time of taking the monthly report.
2) The most necessary thing was the cost of production and the final value that this production will have, in order to estimate the company's profit.
3) Which products are most important to make and why?

To achieve this, the Analytical Production summary was carried out. In this, two graphs were made over time, one with the finished products and another with the processed products. In addition to this, a number of measures, where we can name:

- Production cost.
- Production value.

- That help us estimate the **profit**.
- Finished products.
- Processed products.
- Product categories and product variety.

In addition to this we have two bar charts, one to see where in the product generation the largest number of processed products were found and the product models of these modeled products.

NOTE: The model of the products is like a semi-category, which includes sizes or colors of this product.

In this report each of the problems are met, since the measures present the quantity of products and the price of their production, in addition to an estimated **profit**, which was **43%**.

Added to this, the bar graph that presents the stages of its production (Process Products per Location), allows defining which is the stage in which the products take the longest and this would be **Subassembly**, the reason why this happens could be discovered by going to the site directly, the process is most likely slower and therefore more tends to accumulate.

Depot

In this segment what you have to do is quite simple, what you want is:

1) A small inventory of what we have in the warehouse, what product we are talking about, the quantity it is and the time of year we were at when we had it.

To do this, two tables were generated:

- One presents the number of products that were had at the time the inventory was transferred, normally it was once a year (except for 2014).
- And the other presents the total products that the warehouse received throughout the year and in each month.

In addition to this we have a bar graph that describes the number of products left in the depots.

Sales

This could be one of the most important sections within a company, since it is the one that maintains it monetarily. Here we seek to know different things:

1) The main thing would be the number of sales, orders, sales over time.
2) The best salesperson within the company.
3) Make a relationship between orders and sales.
4) In addition to that we want to know where most of the sales are made.

For this the following was done:

- First, the amount of total sales they have ($109M), their average ($3,5K), the number of orders and the number of customers were indicated (19K). Along with the number of products sold, the average number of products per sale and the number of products in rotation per month.
- Then several graphs were made:

- First we have a sales graph over time, where you can see the evolution of sales. We can highlight that it has an **increasing trend** (dashed line).
- After a map by sales region, where we have three countries in Europe, one for South America called “Central”, the USA was divided into 4 and one for Canada.
- We have the sales, where the largest seller is like **corporate sales**, that is, they were large sales where there was more than one seller involved. Next, we have **Linda Mitchel** and **Jillian Carson**.
- And to relate orders to sales, two scatter graphs were made, one presents this relationship through Regions and the other through clients directly. In the regions we have a positive behavior, the more orders placed, the more sales we obtain and the region with the most sales was **Southwest** in the USA. While for a particular client it is not seen so clearly and normality is lost after 12 orders, but we could conclude that the clients who make larger purchases are between **4 and 12 historical orders**.

Finally, we have the product detail, here you can see each sales characteristic of any product. Such as the orders in which it was included, the quantity, the amount they generated and more data.

HR

In the HR section we want to know how the contracts made and their behavior are divided. For this we carry out:

- Two pie charts, one that divides contracts by management or work areas and another that divides contracts by gender. Here we were able to determine that the area where the most contracts were made was in Manufacturing and the majority of the contracts were for the “M” gender, which would be male.
- We have different metrics that measure the number of contracts (296), staff turnover (2.03%), the number of active and retired workers and the average seniority of active and retired workers.
- A line graph to know the amount over time, this determines that the majority of contracts were made from 2008 to 2010, with its highest peak in December 2008. From mid-2011 onwards, it was normalized to one person every few months.
- Three bar graphs, one dedicated to the age ranges in the contracts, where the vast majority range from 41 to 50 years of age. Another dedicated to management with the best average seniority of active employees, which is led by Quality Assurance and Manufacturing (15.0). And finally one for the average seniority in active workers but now separated by department.
- Finally we have a table with a slightly broader description of the contracts, either for the work performed or for the number of people active in that position.

Customer

Lastly, we have the Custombер area, this is always one of the most important. Here we want to know how our clients are distributed and if they use the promotional email. For this we do the following:

- A pie chart, where it is divided by countries, here you can see that the majority of our clients are in the **USA** (8K).

- Then we have a map that shows how they are distributed in the world, where each point indicates a city. It is notable that the highest density of points is found in the **USA and Europe**.
- And two bar graphs, where one is dedicated to the number of promotional emails we have for each client, which for the majority (56%) **does not have any**. And the other would be from the state in which each client registered, the majority being **California**.
- If we want to dig a little deeper, we have a table that shows the client's complete information, from their email and name to their location.

NEXT STEPS

This report was prepared to fulfill basic monitoring functions in each area of the company, performing a diagnostic analysis. Since in the future, when there is a greater amount of data, it will be possible to carry out predictive analysis. So the next steps would be:

1. Keep this report up to date and continue accumulating data.
2. To later do a predictive analysis.
3. With which to make long-term decisions in different areas.