First step we global\_mega\_sales\_hell\_version.csv changed the file from comma separated values to Microsoft Excel Open XML Format Spreadsheet file

Second step we started cleaning our data

* Removed Unnecessary Columns
* Reordered Columns to make it easier to read data respectively
* Split the dates column by delimiter
* Replace Values to dates columns
* Merge them again
* Change Data Type to Date
* Replaced the values that look inconsistent in each column
* Capitalized the columns that will make it look prettier
* Lowercased Emails so they are all in the same typed form and to look for similarities
* Added New Column for the Total from the true values we got in our table from quantity and price to avoid the misleading and incorrect values in the old column
* Any Email that ends with “@...” will be replaced with “Invalid” by using conditional column using this function

<< if [CustomerEmail] is null then null else if Text.Contains([CustomerEmail], "@...") then "Invalid" else [CustomerEmail] >>

* Any null values were in column named ‘ProductName’ or ‘Category’ replaced with Invalid
* Transformed column named ‘OrderID’ to starting from 1 and filled as a series
* Replaced all values that are invalid in the dates causing errors with nulls
* Added custom column named ‘DaysToRecieve’
* In ‘Currency’ column I used fill down
* Split column named ‘Email’ By Delimiter “@”
* Used all the named in the Splitted column before the delimiter to create new column to fill the invalid and null rows in the column of the ‘CustomerName’
* Removed the old column ‘CustomerName’
* Splitted column ‘CustomerName’ to digit by non-digit
* Deleted the column that contains digits
* Created a function in a new column specifically to check if ‘OrderDate’ is greater than ‘DeliveryDate’ if yes then reverse them else leave it as it was

<< =IF(AND(ISNUMBER(OrderDate),ISNUMBER(DeliveryDate)),IF(DeliveryDate < OrderDate, DeliveryDate, OrderDate), OrderDate) >>

* Created a function in a new column specifically to check if ‘DeliveryDate’ is less than ‘OrderDate’ if yes then reverse them else leave it as it was

<< =IF(AND(ISNUMBER(OrderDate),ISNUMBER(DeliveryDate)),IF(DeliveryDate < OrderDate, OrderDate, DeliveryDate), DeliveryDate) >>

* Removed Old Dates and left the Updated Dates Renamed as the Old Dates
* Copied all the domains in the separated columns and pasted it to csv file
* Copied all the domains in the csv and paste it in an ai tool named “Deep Seek”
* Used it in a function to check for the ‘CustomerEmail’ column

<< =IF(CusomterEmail="","NULL",LET(e, CusomterEmail,d,{"charles.com","hotmail.com","yahoo.com","juarez.com","pratt.com","davis.com","mcdaniel-love.com","newman-hobbs.info","wiggins-gonzales.com","dunn.com","jennings.com","jones-maldonado.com","barnes.com","hernandez.com","massey.net","turner.net","calhoun.info","weber.info","lopez.com","robles.org","jimenez-harris.org","smith.biz","sullivan.com","cooper.com","jones.org","smith.com","curry.com","smith-hall.net","vasquez.net","rodriguez.com","ford.biz","hodges.info","anderson.com","fitzgerald.com","miller-castillo.com","munoz.com","long-simpson.com","lucas-sanders.com","ward.org","murphy-evans.com","smith-estes.com","wright-white.info","gallegos-mann.com","palmer-wood.com","byrd.com","mitchell.com","barker.biz","reeves-salas.com","weeks-alvarado.com","fischer.com","kim-adams.com","perez-morgan.com","bryant.com","wilcox.com","thompson.com","solis-perez.net","armstrong.com","brown.net","armstrong-cannon.biz","li-lawson.com","ramirez.biz","clay-nelson.net","villegas-rodriguez.com","townsend.com","love.com","fitzgerald.com","martin-parker.com","higgins.com","anderson.com","best-sanders.com","bradford-good.com","lewis-murphy.biz","wright.com","castillo.com","mills-mendez.info","reed.org","ford.biz","martin.com","moore-durham.biz","mosley.com","shah.com","martinez.com","stone-warren.com","grant.com","santiago.biz","james.com","garrett.com","wilson-thomas.com","davis.com","serrano-perez.com","davis-moss.com","stewart.org","irwin.com","clark-heath.org","jacobson.info","pierce.com","morgan-powell.com","perez.com","andrade.com","farley.org","nguyen.info","johnson-moore.org","glenn.com","wilkinson-west.com","graham-moore.com","ruiz.com","mcdaniel.com","lee.com","maxwell-johnson.com","ferguson.com","conley-guzman.net","navarro.com","noble-henderson.biz","rose.com","andrade.com","flores.com","white.com","bennett.com","zimmerman-mccarthy.com","richards.com","brooks.net","campbell.com","welch.com","miller-garcia.com","gomez.com","harris-guzman.net","stewart.biz","henry.com","thomas.com","white.com","brady-weber.org","friedman.com","callahan.biz","weber.com"},m,TEXTJOIN("",TRUE,IF(ISNUMBER(SEARCH(d,e)),d,"")),IF(m="",e,IF(ISERROR(SEARCH("@",e)),SUBSTITUTE(e,m,"@"&m),e)))) >>

## If the Email is written with @ missing before the domain it will be added else if the Email the Email is written correctly else if the Email is written as a blank replace it with null else the Email is left as it was

* Replaced null values in ‘CustomerName’ and ‘CustomerEmail’ with unknown
* Filled down Columns ‘UnitPrice’, ‘DeliveryStatus’, ‘PaymentMethod’ ,and ‘Quantity’
* Duplicated column named ‘CustomerName’
* Replaced the ‘space’ in its values with blank
* Lowercased the column named 'CustomerName’
* Replaced the blanks with Unknown
* Added new column by using custom function

<< =IF(AND(C2<>"",C2<>"Invalid",C2<>"Unknown"),C2,IF(AND(D2<>"",D2<>"Invalid",D2<>"Unknown"),LOWER(SUBSTITUTE(D2," ","")&CHOOSE(MOD(ROW(),56)+1,"@gmail.com","@hotmail.com","@charles.com","@yahoo.com","@juarez.com","@jennings.com","@pratt.com","@davis.com","@mcdaniel-love.com","@newman-hobbs.info","@wiggins-gonzales.com","@dunn.com","@mitchell.com","@peterson.com","@jones-maldonado.com","@smith-hall.net","@vasquez.net","@rodriguez.com","@bennett.com","@palmer-wood.com","@rojas.org","@byrd.com","@massey.net","@hernandez.com","@turner.net","@calhoun.info","@weber.info","@rogers.org","@robles.org","@jimenez-harris.org","@smith.biz","@sullivan.com","@cooper.com","@jones.org","@smith.org","@smith.com","@curry.com","@shepard.info","@soto-rush.com","@james-gallagher.com","@wilson.com","@rodgers-harris.info","@carter.net","@howell.com","@anderson.info","@moore.biz","@nicholson-rodriguez.com","@rivers.com","@conley.com","@robinson-robertson.com","@perez.com","@bowen.org","@ochoa.com","@miller-castillo.com","@munoz.com","@anderson.com")), "")) >>

This function work as if CustomerEmail is correct then leave it else if Invalid or Unknown fill it with the CustomerName then add a valid domain randomly from all the domains we have else leave it as it was

* Replaced 0 values in columns ‘UnitPrice’ and ‘Total’ with null
* Replaced Invalid values in column ‘Currency’ with null
* Replaced Unknown values in columns ‘CustomerName’ and ‘CustomerEmail’ with null
* Replaced Invalid values in column ‘CustomerName’ with null
* Used Filtered Rows to keep rows where ‘CustomerName’ does not equal to null
* Replaced Invalid values in columns ‘Category’, City’, ProductName’, and ‘Country’ with null
* Used Filtered Rows to keep rows where ‘City’ does not equal to null
* Used Filtered Rows to keep rows where ‘ProductName’ does not equal to null
* Used Filtered Rows to keep rows where ‘UnitPrice’ does not equal to null
* Changed all data types to be suitable for all columns according to the data
* Started from the data before this step Filled down Columns ‘UnitPrice’, ‘DeliveryStatus’, ‘PaymentMethod’ ,and ‘Quantity’
* Added new column named ‘Delivery Time (Days)’ using this function << Duration.Days([DeliveryDate] - [OrderDate]) >>
* Then added new column named ‘On-Time Delivery ( in 21 Days )’ using this function << if Duration.Days([DeliveryDate] - [OrderDate]) <= 21 then "On-Time" else "Late" >>
* Then added new column named ‘Missing Data’ using this function << if [OrderID] = null or [CustomerName] = null or [CustomerEmail] = null or [Quantity] = null or [UnitPrice] = null then "Missing" else "Complete" >>
* Unlink data from all connections (Optional)
* The data is in its final form cleaned

Third Step Created Pivot tables

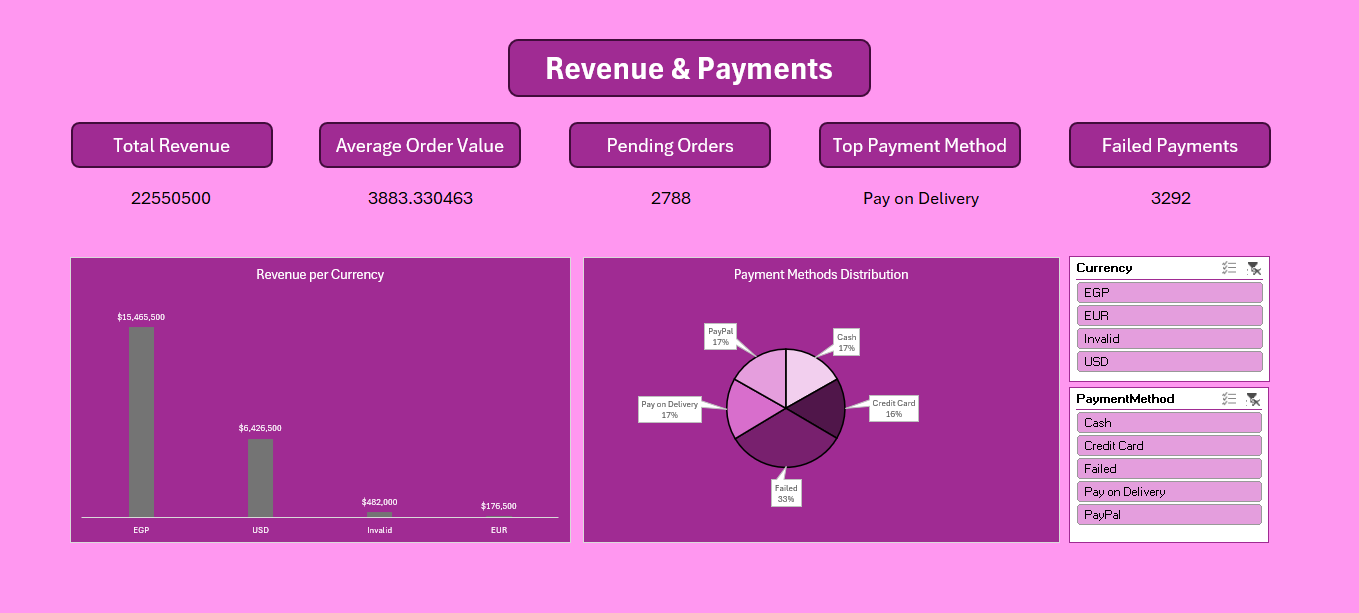
1. DeliveryStatus Count
2. Total Revenue by Currency
3. Payment Methods Distribution
4. Count Valid/Invalid Emails
5. Missing Data
6. Correct Unit Prices & Quantities
7. Sales per Product
8. Sales by Category
9. Orders per Country
10. Revenue per Country
11. City Orders
12. Calculate Avg Revenue per Country
13. Delivery Time
14. Delivery Status
15. Revenue by Payment Method
16. Orders per Payment Method

Fourth step Creating KPIs

1. Total Orders
2. Completed Orders
3. Pending
4. Lost
5. Delivery Rate
6. Total Revenue
7. Average Order Value
8. Top Payment Method
9. Failed Payments
10. Invalid emails
11. Quantity Issues
12. UnitPrice Issues
13. Missing Values (OrderID)
14. Invalid/Missing Delivery Dates
15. Top Product
16. Top Category
17. Total Units Sold
18. Average Price/Product
19. Product Variety
20. Top City by Orders
21. Countries Served
22. Avg Order Value/Country
23. On-Time Rate
24. Pending Deliveries
25. Lost Deliveries
26. Most Used Payment Methods
27. Least Used Payment Methods
28. Failed Transactions
29. Payment Method Share

Fifth step Building Full Interactive Dashboards

1. A screenshot of a computer

   AI-generated content may be incorrect.Order Summary Dashboard
2. Revenue & Payments Dashboard

3. Sales Analysis Dashboard

A blue and white diagram

AI-generated content may be incorrect.

Last step Insights and solved problems

**1.** iPhones are the most profitable product and likely the most demanded.

**2. Payment Method Trends**

**High failure rate (33%)** may indicate payment gateway issues or customer behavior problems.

**3. City-wise Revenue Distribution**

Cairo is the dominant market, suggesting location-based targeting should focus there.

**4. Delivery Performance**

Very low delivery rate (**only 29%**), with nearly **57% orders either pending or canceled** – urgent operational issues.

**5. Order Requests Over Time**

January likely had a **promotion/sales event**; you could replicate this for future growth.

***Thanks to my instructor Ibrahim Elnawassany and mentor Muhammed Adel***