**`1. Handling Missing and Irrelevant Data: Identify and handle missing data in both datasets. Address duplicate entries and irrelevant data points, ensuring data quality.(pr)**

**Findings :**

No Null Values and Duplicate Values were found in the dataset . There were no such irrelevant data found in the dataset**.**

**Conclusion :**

Clean , Reliable & Consistent dataset ready for analysis .

[**Link to Excel Sheet**](https://docs.google.com/spreadsheets/d/1Z45XD7niGMnqv-E2qRmMfhw5gDHcNQLC/edit?usp=sharing&ouid=103377251248088097576&rtpof=true&sd=true)

**2. Name and Email Normalization: Normalize the customer names in Customer.csv to have the first letter capitalized and the rest lower. Standardize the email addresses to follow a consistent format. For example, ensure all email addresses are in lowercase.(pr)**

**Findings :-**

The names were not in proper casing and emails were in all caps . These were normalized the names were turned into proper casing and emails were turned in lower case.

**Conclusion :-**

Customer names & email are now in proper casing.

[**Link to Excel sheet**](https://docs.google.com/spreadsheets/d/1BF6UhFPUDZd-rBgNBShZZcMkXrpeiG6d/edit?usp=sharing&ouid=103377251248088097576&rtpof=true&sd=true)

**3.** Customer-Membership Merge: Merge Customer.csv **with** Membership.csv **using a common key. What key did you use and why?(pr)**

**Findings:-**

Both the Data Were found to have a common key called M\_ID and based on that we merged both dataset . We performed Inner Join .

**Conclusion:-**

We were able to get the Customer & Membership Details in a single sheet.

[**Link to Excel Sheet**](https://docs.google.com/spreadsheets/d/1nuq_UJ1KZAizRkndlc1H4KyYtaivus-q/edit?usp=sharing&ouid=103377251248088097576&rtpof=true&sd=true)

**4. Shipment-Status Integration: Create a new dataset by joining Shipment\_Details.csv with Status.csv. Describe the method used for merging these datasets.(Pi)**

**Findings :-** The data was found to be clean and non redundant. Common key was found SH\_ID and based on that data was joined.

**Conclusion :-**

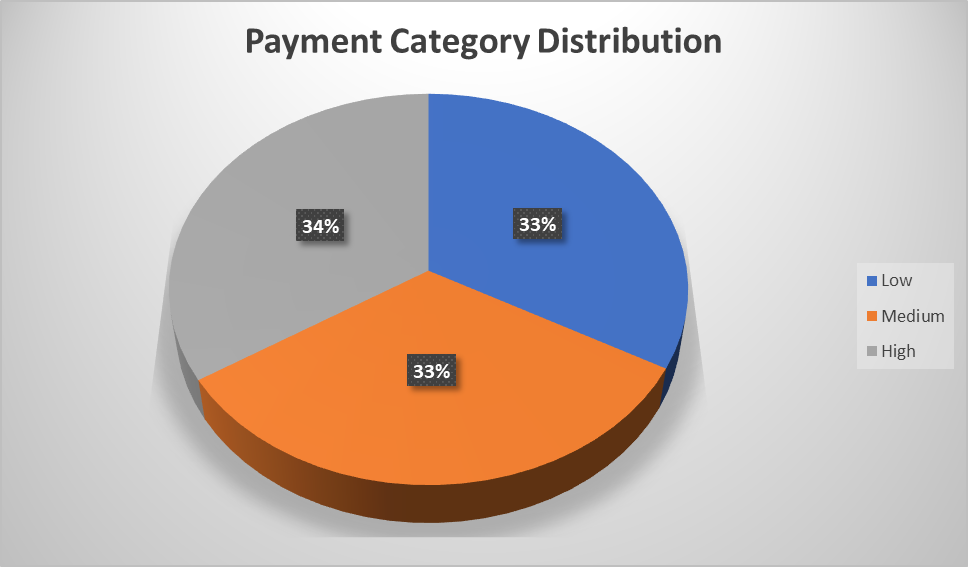
New Dataset was created after joining the Shipment\_Details and Status

**5. Payment Categorization: In Payment\_Details.csv, create a new column that categorizes payments into 'High', 'Medium', and 'Low' based on the amount.(An)**

**Findings :-** The data was found to be Clean and Simple analysis was performed using descriptive statistics and tertile binning, we determined the threshold and based on that the payments were categorized.

**Conclusion :-**

The Dataset was categorized according to the payment amounts . Min value was found to be 312 and Maximum value was found to be 99012.

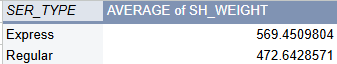
****

[**Link to Excel Sheet**](https://docs.google.com/spreadsheets/d/1m1f_MnizctEuXvGk_5IBEDSNh7Csjn5E/edit?usp=drive_link&ouid=103880064888448540547&rtpof=true&sd=true)

**6.** Average Shipment Weight: Calculate the average shipment weight for each type of service in Shipment\_Details.csv**.(Pi)**

**Findings :- Using pivot table Functionality the average shipment weight for each service type was calculated.**

**Conclusion:-**

****

[**Link to Excel Sheet**](https://docs.google.com/spreadsheets/d/1VchzJIpV28hfmIDUkI0dagGAInmJVjml/edit?usp=sharing&ouid=104200202458993417106&rtpof=true&sd=true)

**7. Membership Duration Calculation: Use a date function to find out how long each customer has been a member (from Membership.csv).(Pr)**

**Findings:-** Dataset was clean and no redundancy found. Using the Start date and end date Column we found out the no of days the customer has been a member.

**Conclusion :-**

We found how many number of days the person were a member.

At an average members were subscribed for 4000 days .

[**Link to Excel Sheet**](https://docs.google.com/spreadsheets/d/1l9kqiUizachvD3ol9t441VUc4liACMZ6/edit?usp=sharing&ouid=103377251248088097576&rtpof=true&sd=true)

**8. Total Customer Payment: Create a formula to calculate the total amount paid by each customer in Payment\_Details.csv.(An)**

**Findings :- We found that each customer id was unique in the dataset.**

**For each customer we calculated the total amount by using sumifs if the payment status is paid.**

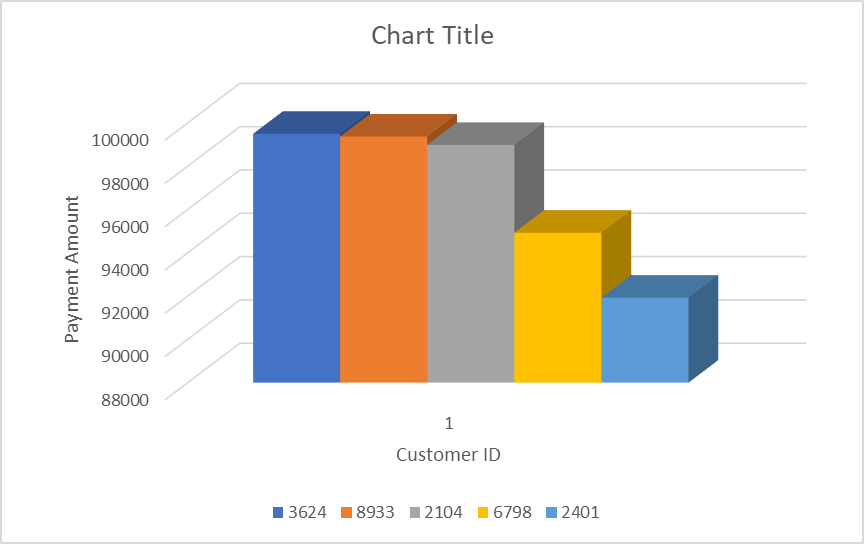
**Conclusion :- Total amount paid by each customer averages around 25000.There are around 100 customers who haven’t made any payment.**

[**Link to the Excel Sheet**](https://docs.google.com/spreadsheets/d/1m1f_MnizctEuXvGk_5IBEDSNh7Csjn5E/edit?usp=drive_link&ouid=103880064888448540547&rtpof=true&sd=true)

**9. Top Customers Analysis: Identify the top 5 customers by payment amount in Payment\_Details.csv.(An)**

**Findings:-** The top 5 customers by payment amount was found by sorting the entire dataset on the Payment amount and then taking the top 5 records.

**Conclusion :-** Highest payment made by the customer**:** 99492

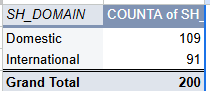
****

[**Link to the Excel Sheet**](https://docs.google.com/spreadsheets/d/1m1f_MnizctEuXvGk_5IBEDSNh7Csjn5E/edit?usp=drive_link&ouid=103880064888448540547&rtpof=true&sd=true)

**10. Shipment Domain Comparison: Analyze the Shipment\_Details.csv dataset to find out which shipment domain (Domestic/International) has more shipments.(Pi)**

**Findings:- The Dataset was Clean & No Redundancy was found. Using the Pivot table the Shipment Count was found based on their domains.**

**Conclusion:-**

****

[**Link to dataset**](https://docs.google.com/spreadsheets/d/1A4VJOLk-kAHIpKGR1-VLJ7ip_fu2kUmY/edit?usp=sharing&ouid=116015531523725686088&rtpof=true&sd=true)

**11. Top Employees Characterization: What are the common characteristics of the top 5 employees managing the highest number of shipments?(Sh)**

Analyzed Both Employee\_Details and Employee\_manages\_shipments.csv

**Findings:-**

| **Top Employee ID** | **Employee Name** | **Designation** | **Branch** | **Shipment\_Count** |
| --- | --- | --- | --- | --- |
| 2 | Zoya | Transport manager | TN | 1 |
| 7 | Guy | Transport manager | TN | 1 |
| 11 | Rita | Manager | WI | 1 |
| 18 | Jennie | Project director | WA | 1 |
| 23 | Yasmeen | In House logistics executive | TX | 1 |

**Conclusion:**

From the analysis, all five 'top' employees (Zoya, Guy, Rita, Jennie, and Yasmeen) each managed a single shipment. This indicates that in this dataset, 'top' refers to any employee who has managed at least one shipment, rather than a distinction based on higher volume.

Regarding common characteristics, there is no single predominant designation or branch among these five individuals. However:

* Designation: 'Transport manager' is the most frequent designation, accounting for two out of the five employees (Zoya and Guy).
* Branch: The 'TN' branch is the most common location, also representing two out of the five employees (Zoya and Guy).

The remaining employees hold diverse roles like 'Manager', 'Project director', and 'In House logistics executive' across different branches (WI, WA, TX)."

[Link to the Excel sheet](https://docs.google.com/spreadsheets/d/1X9zZsEn55zejGFKeI_tJscoLvY_I-tJe/edit?usp=sharing&ouid=104200202458993417106&rtpof=true&sd=true)

**12. Conditional Formatting for Weight: Use conditional formatting in Shipment\_Details.csv to highlight shipments weighing over 500 units.(Pr)**

**Findings:-** By using the Conditional Formatiting we have Highlighted the Shipments Weighing over 500 units. “Formats only cells that Contain”

**Conclusion :-** Shipments with Weight Greater than 500 units are now automatically highlighted Making the Visibility of the Document Better.

[**Link to the Excel Sheet**](https://docs.google.com/spreadsheets/d/1A4VJOLk-kAHIpKGR1-VLJ7ip_fu2kUmY/edit?usp=sharing&ouid=116015531523725686088&rtpof=true&sd=true)

**13. Employee Detail Dropdown: Create a dynamic drop-down list in Employee\_Details.csv to select and display details of an employee.(Sh)**

**Dynamic Table in the Excel Sheet :-**

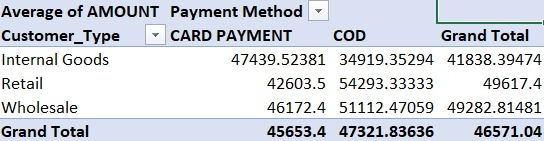
****

[**Link to the Excel Sheet**](https://docs.google.com/spreadsheets/d/1X9zZsEn55zejGFKeI_tJscoLvY_I-tJe/edit?usp=sharing&ouid=104200202458993417106&rtpof=true&sd=true)

**14. Payment Summary Pivot Table: Implement a pivot table to summarize the average payment amount by customer type and payment mode.**

**Findings :- Using the VLOOKUP Function and Pivot Table Functionality of Excel data was grouped on customer type and payment mode and based on that average payment was calculated. (An)**

**Conclusion :- Retail category has the lowest total payment amt.**

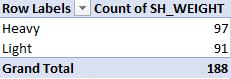
****

[**Link to the Excel Sheet**](https://docs.google.com/spreadsheets/d/1m1f_MnizctEuXvGk_5IBEDSNh7Csjn5E/edit?usp=drive_link&ouid=103880064888448540547&rtpof=true&sd=true)

**15. Shipment Weight Categorization: Categorize shipments in Shipment\_Details.csv as 'Heavy' or 'Light' based on weight.(Pi)**

**Findings:-** A new Column called Weight\_Category was introduced and based on a threshold calculated by utilizing the descriptive statistics of the column SH\_WEIGHT we categorized the diff weights into Heavy or Light.

**Conclusion :-** After analyzing the Weight Categories we found out 97 out of 188 rows are heavy and the rest of all are light weight .

****

[**Link to the Excel Sheet.**](https://docs.google.com/spreadsheets/d/1A4VJOLk-kAHIpKGR1-VLJ7ip_fu2kUmY/edit?usp=sharing&ouid=116015531523725686088&rtpof=true&sd=true)

**16.** Employee Shipment Count: In employee\_manages\_shipment.csv**, use an array formula to count the number of shipments managed by each employee.(Sh)**

**Findings:-** A new Sheet was created and two columns were added and by using the query formula all the columns were populated.

Query Formula used :-   
=QUERY(Shipment\_Log!A:C, "SELECT A, COUNT(B) GROUP BY A LABEL COUNT(B) 'Shipment Count'", 1)

[**Link to the Excel**](https://docs.google.com/spreadsheets/d/1X9zZsEn55zejGFKeI_tJscoLvY_I-tJe/edit?usp=sharing&ouid=104200202458993417106&rtpof=true&sd=true)

[**Sheet**](https://docs.google.com/spreadsheets/d/1X9zZsEn55zejGFKeI_tJscoLvY_I-tJe/edit?usp=sharing&ouid=104200202458993417106&rtpof=true&sd=true)

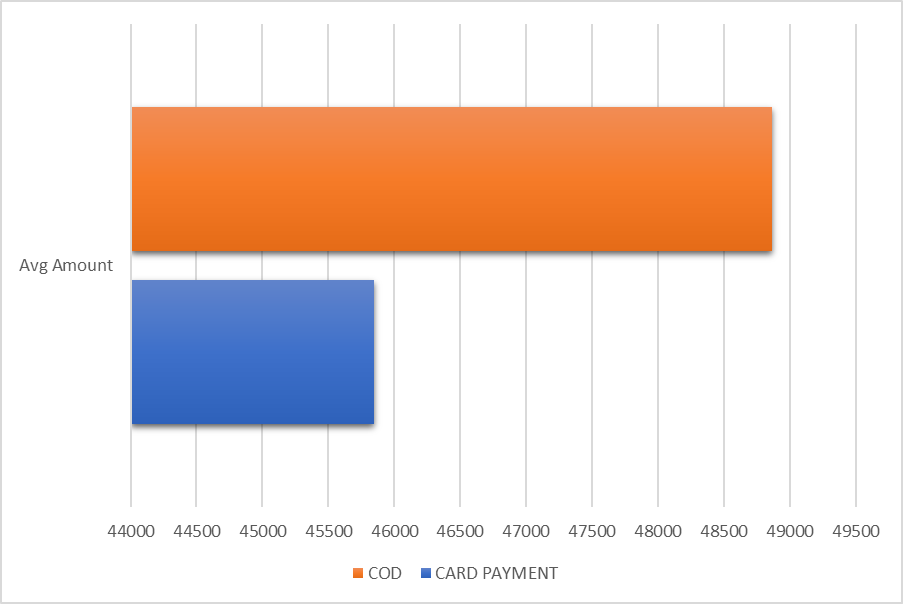
**17. Dynamic Payment Analysis: In Payment\_Details.csv, create a dynamic named range for each payment mode (e.g., CARD PAYMENT, COD). Then use a formula to calculate the average payment amount for each mode(An)**

**Findings:-** We can define our dynamic range using excel inbuilt function Name Manger.

Using the filter() function we found out the amounts for both the Payment Modes: Cash On Delivery (COD) and Card Payment. Applied AVERAGE() on the data fetched using our dynamic range

Conclusion: The Average Amount for Card Payment : 45,848

The Average Amount for Card Payment : 48,862

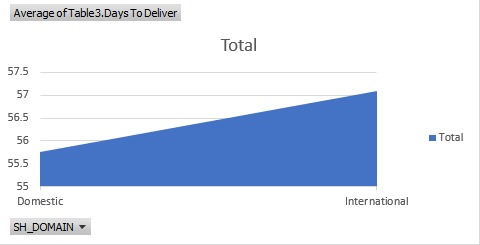


[**Link to the Excel Sheet**](https://docs.google.com/spreadsheets/d/1m1f_MnizctEuXvGk_5IBEDSNh7Csjn5E/edit?usp=drive_link&ouid=103880064888448540547&rtpof=true&sd=true)

**18. Average Delivery Time Analysis: Analyze the combined data of Shipment\_Details.csv and Status.csv to determine the average delivery time for shipments.(Consider using the DATEDIF function or similar.) How does the average delivery time vary between domestic and international shipments?(Pr)**

**Findings:-** WE have combined the Data of Shipment\_Details and Status based on the Common key SH\_ID to determine the average delivery time for shipments using the DATEDIF function. Using the pivot function we have analyzed the domestic and international shipments difference.

**Conclusion :-**

****

[**Link to the Excel Sheet**](https://docs.google.com/spreadsheets/d/1A4VJOLk-kAHIpKGR1-VLJ7ip_fu2kUmY/edit?usp=sharing&ouid=116015531523725686088&rtpof=true&sd=true)

**19.) Designation-wise Shipment Charge: Using a pivot table, analyze the employee\_manages\_shipment.csv and Shipment\_Details.csv datasets to find out which employee designation has the highest average shipment charges. Incorporate slicers or timeline filters for dynamic data exploration.(Sh)**

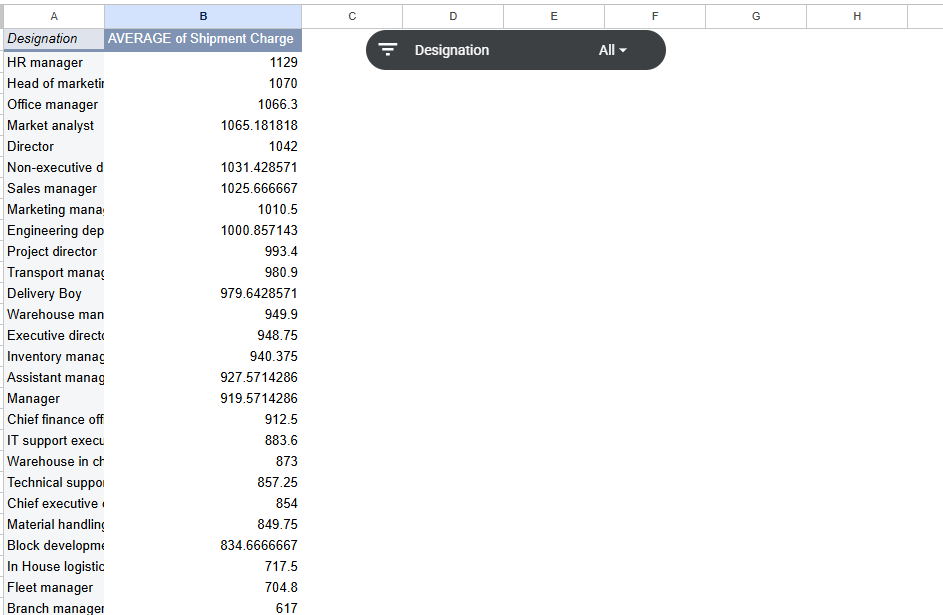
Sheets Analyzed :-

1. Shipment\_Details.csv
2. Employee\_Details.csv
3. Employee manages sheepment.csv

Steps :-

1. Relevant information from all three sheets were clubbed together
2. Using PIvot table functionality the Average Charges Designation wise was calculated.
3. Added Slicer using Slicer Functionality.

**Results :-**

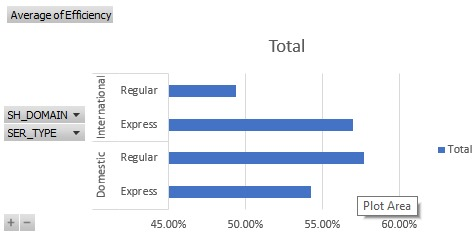
****

[**Excel Sheet Link**](https://docs.google.com/spreadsheets/d/1sIv9EWG_DFks5ew1YZLSWYDi1Rw1048V/edit?usp=sharing&ouid=104200202458993417106&rtpof=true&sd=true)

**20.) Shipment Efficiency Analysis: Analyze the Shipment\_Details.csv to calculate the efficiency of shipments. Define efficiency as the ratio of SH\_WEIGHT to SH\_CHARGES, categorized by SH\_DOMAIN (Domestic/International) and SER\_TYPE (Regular/Express). Identify the most and least efficient categories.(Pi)**

**Findings:-** Analyzed the Shipment\_Details.csv dataset and created a new column called efficiency by sh\_weight/sh\_charges .

**Conclusion :-** The most efficient category is Domestic Regular with 57.71% and least efficient is international Regular with 48.40% .

****

[**Link to the Excel Sheet**](https://docs.google.com/spreadsheets/d/1A4VJOLk-kAHIpKGR1-VLJ7ip_fu2kUmY/edit?usp=sharing&ouid=116015531523725686088&rtpof=true&sd=true)

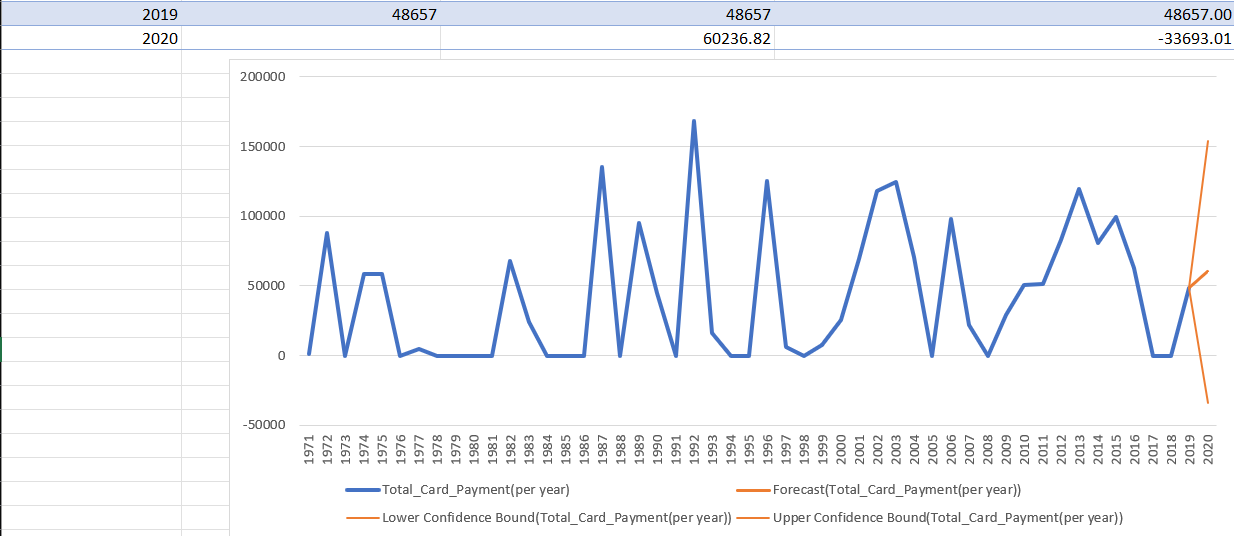
**21. Payment Forecast: Using the Payment\_Details.csv, create a time-series forecast to predict future payment amounts for the next year based on historical data. Break down the forecast by payment mode. (Hint: Utilize Excel's forecasting tools, like the Forecast Sheet feature. Ensure the data is in a chronological order and consider using a line chart for visualization. Analyze trends and seasonality in the payment data.)(An)**

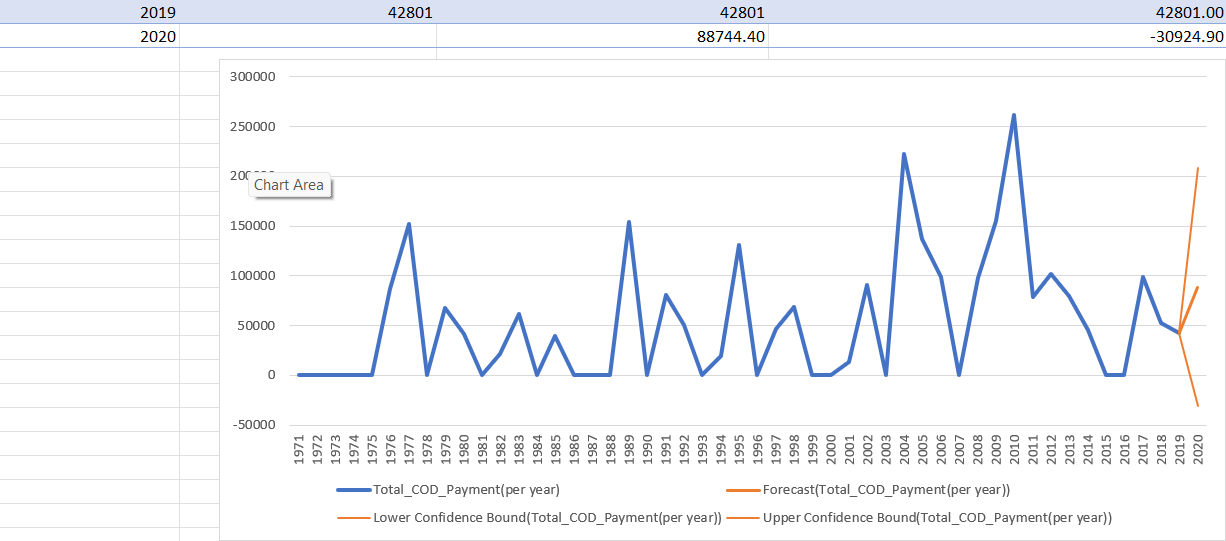
**Findings:-** We found that our date records must be evenly distributed for utilizing Excel’s Forecast Sheet feature so accordingly we calculated the total amount spent each year (1971 - 2019) and applied forecasting .

**Conclusion:** Based on given historical data,Card payments shows fluctuating behavior, but the trend becomes smoother and more upwards from 2000–2015.

Forecasted Amount(Card Payment) for the next year i.e.2020: 60236.82

Forecasted Payment(C.O.D) Amount for the next year i.e.2020: 88744.40

****

****

[**Link to Excel Sheet**](https://docs.google.com/spreadsheets/d/1m1f_MnizctEuXvGk_5IBEDSNh7Csjn5E/edit?usp=drive_link&ouid=103880064888448540547&rtpof=true&sd=true)

**22.) Customer Segmentation: Segment customers in Customer.csv based on their transaction behavior from Payment\_Details.csv and membership duration from Membership.csv. Create segments such as 'High Value - Long Term', 'Low Value - Short Term', etc., based on total payment amounts and membership lengths.(Sh)**

Analyzed Files :-

1. Customer.csv
2. Payment\_Details.csv
3. Membership.csv

Steps:-

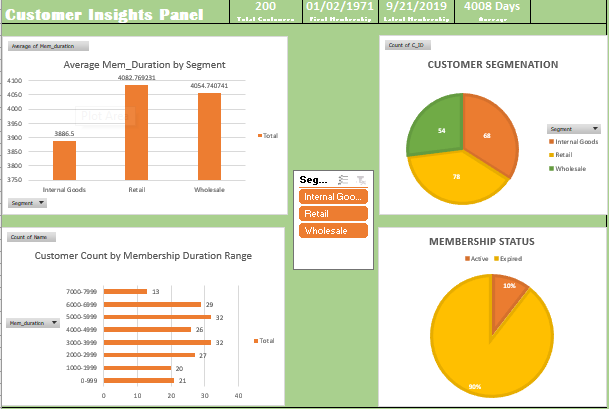
1. Membership id and Days since Membership Column was taken.
2. Customer id was derived from membership id
3. Based on the Customerid the Total amount paid by that customer was taken .
4. Total amount and Days since Membership Columns were analyzed by mean median and from that we selected the threshold value.
5. Based on The threshold value Customer Segmentation was done.

[**Link to the excel file**](https://docs.google.com/spreadsheets/d/1BEOCFE9NoFln5spAu19Od92NJHSTvqHW/edit?usp=sharing&ouid=104200202458993417106&rtpof=true&sd=true)

**Part 2**

1. **Customer Insights Panel:**

Displayed key metrics like total number of customers, customer segmentation (e.g., retail, wholesale), and membership status distribution.

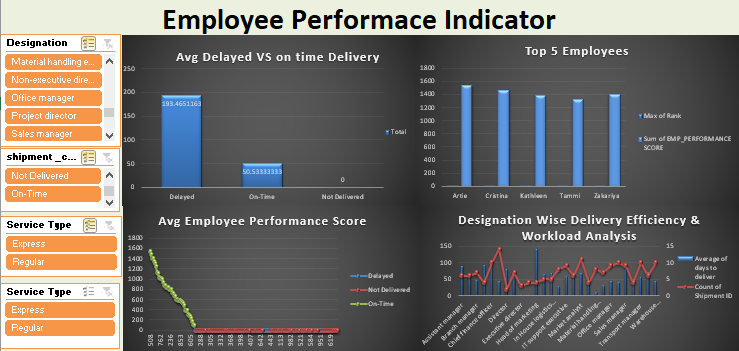


**Insights :**

* The average membership duration for retail customers is highest with 4082 days and lowest is internal goods is 3886.5.
* The customers are segmented uniformly into internal goods retail wholesale with retail having maximum customers (78).
* Most of the customers stay between 5000 to 6000 days .
* Only 10% of the memberships are active and rest of 90% are expired .
* The first membership was on the date 2nd jan 1971 and the latest membership was on 21st sep 2019.

**2. Employee Performance Tracker:**

Included metrics such as number of shipments handled, average handling time, and a ranking of employees based on their shipment management efficiency.

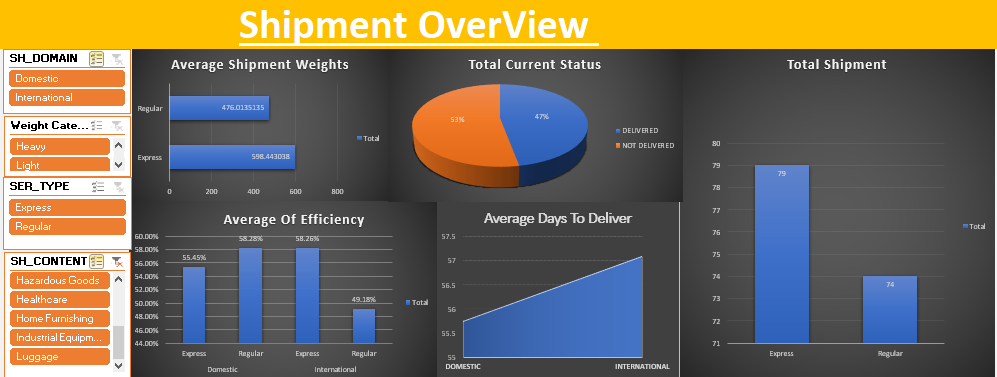
****

**Insights :**

* According to this data , we found that the Delayed shipments are taking unsurmountable amount of time more than On-Time shipments .
* According to this dashboard we can see the top 5 employees based on employee performance score with the top performer having 1546.27 points .
* According to this Dashboard we can see the designation that handled most shipments is “Head Of Marketing” with the value 141 And the least shipments are handled by “Executive Director”.

3. **Shipment Overview**:

Visualized total shipments, average shipment weight, efficiency (weight vs. charges), and status (delivered, not delivered) with filters for domestic and international shipments.

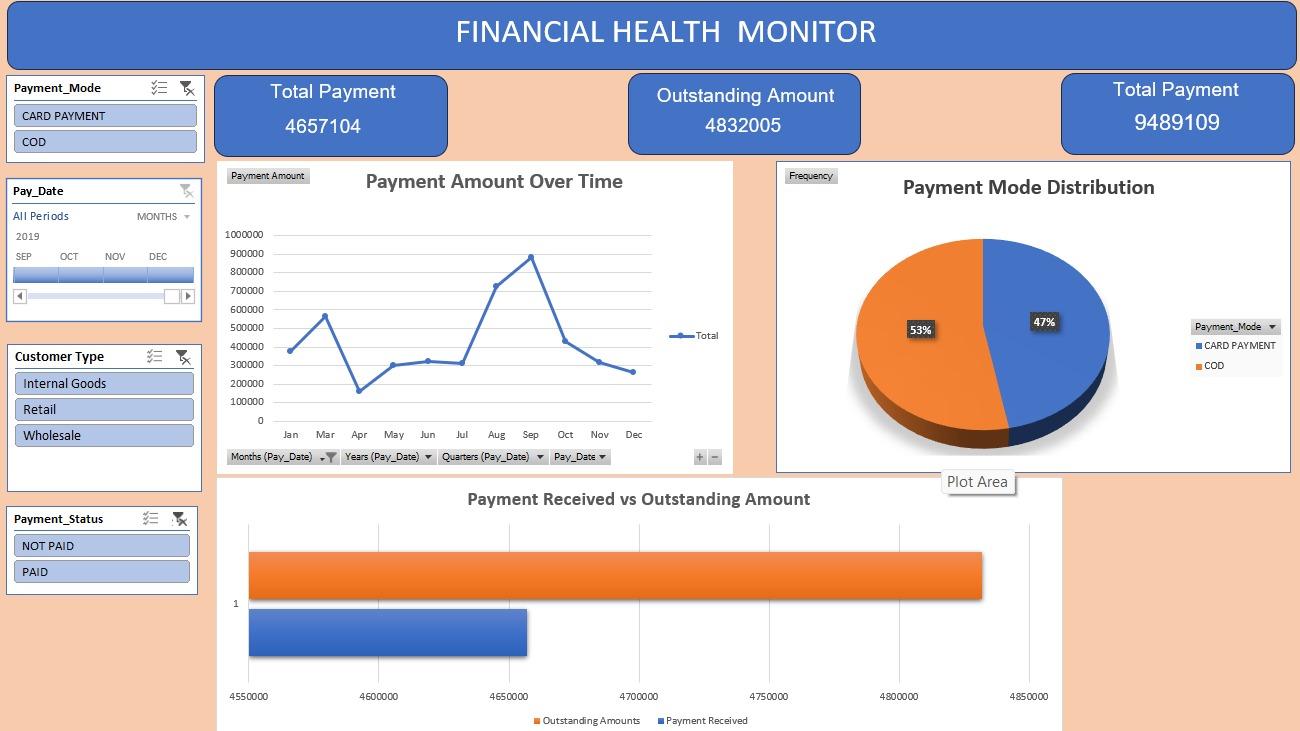


**Insights :**

* According to this dashboard we can simply see the shipment weights of shipment type express has (598 kgs) which is more then regular (476 kgs ).
* In overall shipments 53% were not delivered and the rest of 47% were delivered .
* According to this data most of the shipments are express so we can say customers prefer express shipments to get their orders delivered .
* According to this data ,in the domestic field express shipments are more efficient than regular . And in the international field express shipments are more efficient than regular.

1. **Financial Health Monitor**:

Create a financial summary showing total payments received, outstanding amounts, payment mode distribution, and a monthly trend of payments.



**Insights :**

* According to the line chart we can observe high fluctuations in monthly payment over payments made since 1971 till 2019. Peak value can be observed in August followed by a decline post September.
* The pie chart reflects the distribution of total payment made by each mode. Both Cash on Delivery(53%) and Card Payment(47%)
* Indicating that customers are evenly divided in their payment mode preference.
* The Outstanding Amount stands at 4,832,005, which is slightly more than the Payment Received (4,657,104). Highlighting a cash flow concern, as over 50% of total dues remains unpaid.
* The Total Payment value of is 9,489,109 inclusive of both received and outstanding amounts.