



DATA ANALYSIS FOR BUSINESS-1

Project report



MEMBERS:

HAROON SALEEM(25F-5001)

HAIDER HABIB(25F-5027)

QASIM ZAHID(25F-5004)

DANIA ATHAR(25F-5012)

Contents

1. Problem Statement.....	2
1.1 Why it matters:	2
2 Objective	3
2.1 Descriptive Objective: Financial Transparency	3
2.2 Diagnostic Objective: Performance and Efficiency	4
2.3 Predictive/Inferential Objective: Risk Management and Forecasting.....	4
3. Project description:	4
3.1 Key Findings from the Shape.....	Error! Bookmark not defined.
3.2 Why This Matters for our Project	Error! Bookmark not defined.
this chart focuses on the frequency and spread of the actual monetary values	Error! Bookmark not defined.
3.3 Central Tendency (The Peak)	Error! Bookmark not defined.
3.4 The "Skew" in your Data	Error! Bookmark not defined.
3.5 Standard Deviation and Spread.....	Error! Bookmark not defined.
3.6 Connection to your Z-Scores	Error! Bookmark not defined.
4.Z-Score for Income	9
5.What do these tables show?.....	Error! Bookmark not defined.
6. Z-Score Expense Table (The Risk Forecast)	16
6.1 Why are these Z-Scores so important?	16
6.2 Why these tables matter for your report:.....	18
The Prediction (What should we expect?)	20
Who is responsible?	20
7.Binomial distributions.....	20
8.What is this telling us?	22
9.Poisson distribution.....	23
Conclusion:	24

1. Problem Statement

The core problem addressed by this project is the **lack of transparency and predictability in the operational efficiency and financial health of the travel service firm**. While the business is generating a healthy net profit (£4,180.49), there are underlying inefficiencies and uncertainties:

1. **Agent Performance Disparity:** Significant variance exists between agents (e.g., Sara generating £1,494 in profit vs. Ali Haider resulting in a £300 loss), suggesting that revenue generation is not uniform across the team.
2. **Cost Management:** High-cost categories like "Flight Costs" and "Appointment Costs" are impacting margins, particularly for specific agents, without a clear strategy for cost containment.
3. **Financial Predictability:** Without statistical validation (like the Z-scores and Probability distributions calculated), the business cannot accurately forecast future cash flows or understand the "true" average value of a transaction versus an outlier.

1.1 Why it matters:

Understanding these factors is critical for the long-term sustainability and growth of the business for several reasons:

- **Strategic Decision Making:** By identifying that Ali Haider's transactions are resulting in losses due to flight costs, management can intervene—either by retraining the agent, renegotiating flight rates, or adjusting the service fee structure.

- **Risk Mitigation (Confidence Intervals):** The Z-score analysis provides a 95% and 99% confidence interval. This matters because it tells the business what "normal" income looks like (£62.47 to £77.40 per transaction). If future averages fall below this, management knows immediately that something is wrong (e.g., a market shift or internal error) before a crisis occurs.
- **Scaling and Growth (ROAS):** The data reveals a highly efficient marketing spend (14.6x Return on Ad Spend). Recognizing this allows the business to confidently increase its "Ads Expense" budget to scale revenue, knowing that the infrastructure (led by agents like Sara and Hamza) can convert that spend into profit.
- **Operational Optimization:** The probability analysis (59% chance of income) allows for better resource planning. For instance, knowing the probability of specific agents handling certain income types (Conditional Probability) helps in assigning leads to the most effective team members.

2 Objective

2.1 Descriptive Objective: Financial Transparency

The immediate goal is to accurately track and categorize every pound entering and leaving the business.

- **Goal:** To determine the exact Net Profit for the month (£4,180.49).
- **Goal:** To visualize cash flow and identify where the money is going (e.g., categorizing expenses into Ads, Flights, and Appointments).

2.2 Diagnostic Objective: Performance and Efficiency

The project aims to look "under the hood" to see why certain results are occurring.

- **Agent Analysis:** To evaluate the contribution of each agent. This objective identifies that while some agents are highly profitable (Sara and Hamza), others are operating at a loss (Ali Haider), allowing for targeted management intervention.
- **Cost-Benefit Analysis:** To measure the effectiveness of marketing. For example, the objective was to see if the £500 spent on ads was justified by the £7,343 in income (which resulted in a high 14.6x return).

2.3 Predictive/Inferential Objective: Risk Management and Forecasting

This is the most advanced part of the project, represented by the "Z-score" and "Probability" sheets.

- **Statistical Validation:** To use **Confidence Intervals** to establish a "baseline" for the business. The objective here is to define what a "normal" transaction looks like so the company can identify outliers or future declines early.
- **Probability Modelling:** To calculate the likelihood of future success. For instance, determining the 59% probability of a transaction being income helps the business project its cash flow and staffing needs for the following month.

3. Project description:

This project is a real data set of a company. The project clearly shows dataset analyzed in this project consists of a comprehensive financial ledger documenting

177 individual transactions comprising 105 income entries and 72 expense records captured during the operational period of April 2026 for a travel and visa service firm.

Data Variables and it's relevance

The given dataset is used to create a complete Transactional Ledger that will be an account of the transactions carried out by the consultancy firm in the month of April 2026. By considering various aspects of each transaction right from identifying the clients, agents, and services provided and the banks where money was deposited, the ledger will give a detailed insight into the company's sources of income and expenses. In terms of conducting a research study, this data will be considered a primary source that helps in determining the state of the business, productivity of agents, and market demands. Below is a description of each of the variables given in the dataset.

Date and Day of the Week

These variables show the order of the transactions with regard to the timing of the transactions, which provides both a time stamp and the day in the week when the transaction was made. Such information enables the tracking of all the business activities taking place during the month of April 2026.

In a research context, these serve as the foundation for **Time-Series Analysis**. They enable the researcher to identify cyclical trends, peak activity periods, and seasonal fluctuations. For instance, analyzing the "Day of the Week" helps determine if business volume is concentrated on specific days (like the high frequency of entries on Fridays), which is vital for operational forecasting and workforce scheduling.

Customer and agent name

These categorical variables represent the main actors within each transaction, namely the client who receives the service as well as the internal agent (Agent) in charge of the case file. There is thus a clear association between the origin of income and the people offering services.

These two variables are vital to Behavioral and Performance Modeling because customer names help in conducting "Cohort Analysis," which will help in determining whether the customers who used their services before belong to the current cohort or not. "Productivity Assessment" can be conducted by using agent names because it is possible to find out the contribution of each agent to total output.

Amount

This is the fundamental numerical factor that expresses the dollar value of each transaction. It acts as the fundamental measurement standard for each transaction's size, varying from smaller insurance costs to larger airplane costs.

Column "Amount" forms the Dependent Variable for all types of statistical testing. This makes it possible to calculate the following Descriptive Statistics parameters: the mean value, median, and standard deviation of the cash flow of the firm. In a business report, this information serves as a parameter for determining the financial status of the firm as well as conducting "Variance Analysis."

Bank

This variable is used to distinguish which financial institute (Lloyds or Anna) was used to conduct this particular transaction. It helps identify the place where money was received from or spent.

In the analysis of Corporate Finance and Risk Management, the measure used is referred to as "Liquidity Distribution." It is useful in establishing whether the organization depends on any particular financial partner and also ensures that the audits are carried out accurately. The study of banks could also show preference for quick payments or low fees.

Type and Category

These factors give qualitative background to every transaction; "Type" describes the kind of service that is being considered (such as Visa Cost, First, and Flight Income), whereas "Category" classifies every item as either "Income" or "Expense."

These are critical for Segmental Analysis. Through the sorting of the information in terms of "Type," it becomes possible to distinguish the lines of business that bring in higher profits, such as whether Immigration or Travel operations are more profitable. "Category" serves as the principal criteria for the calculation of the "Net Profit Margin."

4 Results

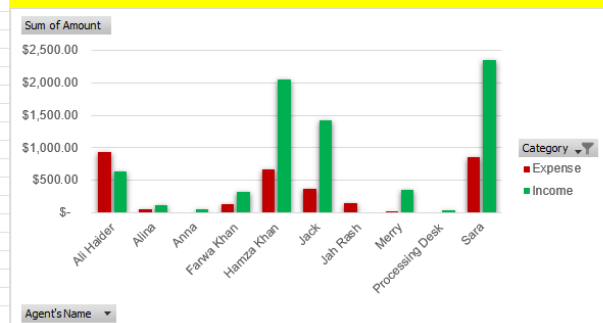
4.1 Financial Performance Analysis by Agent

Profit (loss) Table			Income and Expenses Table			
Sum of Amount	Column Labels		Sum of Amo	Column Labels		
Row Labels	Profit (loss)		Row Labels	Expense	Income	
Ali Haider	\$ (300.76)		Ali Haider	\$ 939.76	\$ 639.00	
Alina	\$ 62.77		Alina	\$ 47.23	\$ 110.00	
Anna	\$ 50.00		Anna	\$	\$ 50.00	
Farwa Khan	\$ 192.24		Farwa Khan	\$ 122.76	\$ 315.00	
Hamza Khan	\$ 1,398.95		Hamza Khan	\$ 661.05	\$ 2,060.00	
Jack	\$ 1,057.16		Jack	\$ 371.84	\$ 1,429.00	
Jah Rash	\$ (143.00)		Jah Rash	\$ 143.00		
Merry	\$ 334.00		Merry	\$ 16.00	\$ 350.00	
Processing Desk	\$ 35.00		Processing Desk		\$ 35.00	
Sara	\$ 1,494.13		Sara	\$ 860.87	\$ 2,355.00	
Grand Total	4180.49		Grand Total	\$ 3,162.51	\$ 7,343.00	

Profit (loss) report by agent chart



Income and Expenses by agent Chart

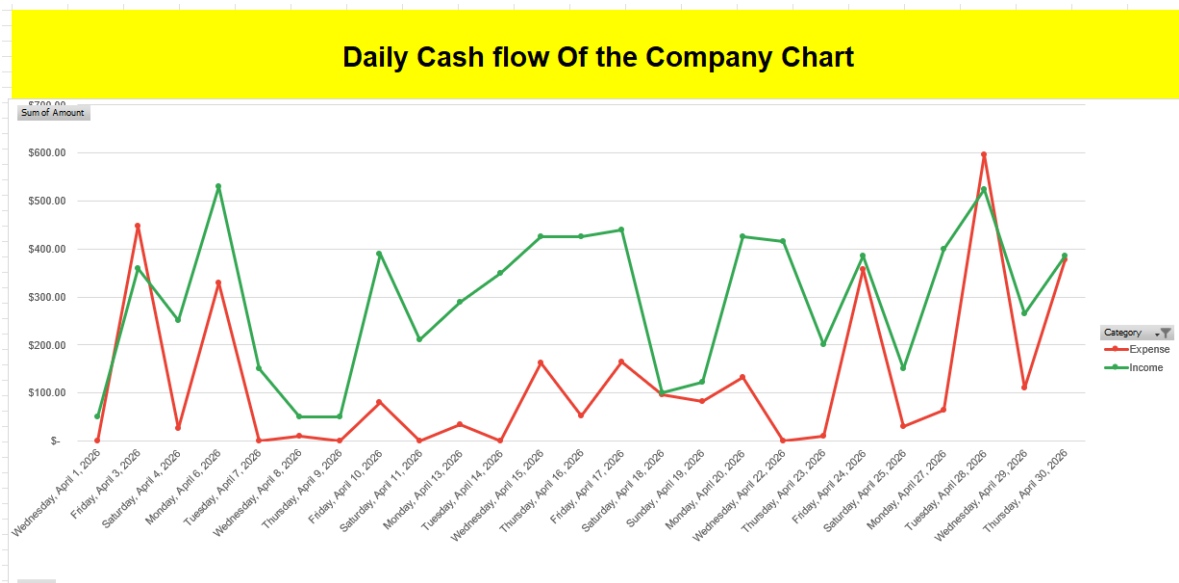


From the data presented above, it can be observed that the organization is making profits because its Total Net Profit stands at \$4,180.49, which is the result of its gross income of \$7,343.00 minus expenses of \$3,162.51. As it pertains to the performance aspect, it seems that it is dominated by the high-performing individuals within the group, particularly Sara and Hamza Khan, as they contribute more than 69% to the total profit made by the organization. Even as some of the members make good profit margins, the "Income and Expenses Table" demonstrates that some of the individuals are performing poorly because their income and expense margins are significantly disproportionate.

As indicated visually in the "Profit (loss) report by agent chart," the high level of profitability is clearly depicted by very high spikes and very clear negatives for those that have a loss. The second chart known as the "Income and Expenses by agent Chart" clearly shows how efficient agents like Jack and Hamza are with their operations because their income spikes are much larger compared to their expense's spikes. On the other hand, organizations such as Processing Desk and Anna barely generate any financial activity, thus meaning they provide support functions within the organization.

4.2 Financial Operations Report

Daily Cashflow Table			
Sum of Amount	Column Labels		
Row Labels	Expense	Income	
Wednesday, April 1, 2026	\$ -	\$ 50.00	
Friday, April 3, 2026	\$ 447.14	\$ 360.00	
Saturday, April 4, 2026	\$ 25.50	\$ 250.00	
Monday, April 6, 2026	\$ 328.46	\$ 531.00	
Tuesday, April 7, 2026	\$ -	\$ 150.00	
Wednesday, April 8, 2026	\$ 10.50	\$ 50.00	
Thursday, April 9, 2026	\$ -	\$ 50.00	
Friday, April 10, 2026	\$ 80.66	\$ 390.00	
Saturday, April 11, 2026	\$ -	\$ 210.00	
Monday, April 13, 2026	\$ 34.56	\$ 290.00	
Tuesday, April 14, 2026	\$ -	\$ 350.00	
Wednesday, April 15, 2026	\$ 162.32	\$ 425.00	
Thursday, April 16, 2026	\$ 51.99	\$ 425.00	
Friday, April 17, 2026	\$ 164.23	\$ 440.00	
Saturday, April 18, 2026	\$ 95.74	\$ 100.00	
Sunday, April 19, 2026	\$ 83.00	\$ 123.00	
Monday, April 20, 2026	\$ 131.94	\$ 425.00	
Wednesday, April 22, 2026	\$ -	\$ 415.00	
Thursday, April 23, 2026	\$ 10.50	\$ 200.00	
Friday, April 24, 2026	\$ 357.28	\$ 385.00	
Saturday, April 25, 2026	\$ 29.51	\$ 150.00	
Monday, April 27, 2026	\$ 63.82	\$ 399.00	
Tuesday, April 28, 2026	\$ 596.64	\$ 525.00	
Wednesday, April 29, 2026	\$ 110.29	\$ 265.00	
Thursday, April 30, 2026	\$ 378.43	\$ 385.00	
Grand Total	\$ 3,162.51	\$ 7,343.00	



"The Daily Cashflow Table," along with its line graph, shows one month's worth of cash flow that is both volatile but mainly positive, marked by regular instances of income peaks and careful expenditure control. Through all of April 2026, there were consistent profits recorded, as shown by the fact that the green "Income" line always stayed above the red "Expenses" line during the bulk of the reporting month. Income peaks include those seen on April 6 (\$531.00) and April 28 (\$525.00); this ensured sufficient liquidity to cover expensive days, including April 28, when the monthly peak of expenses (\$596.64) was recorded.

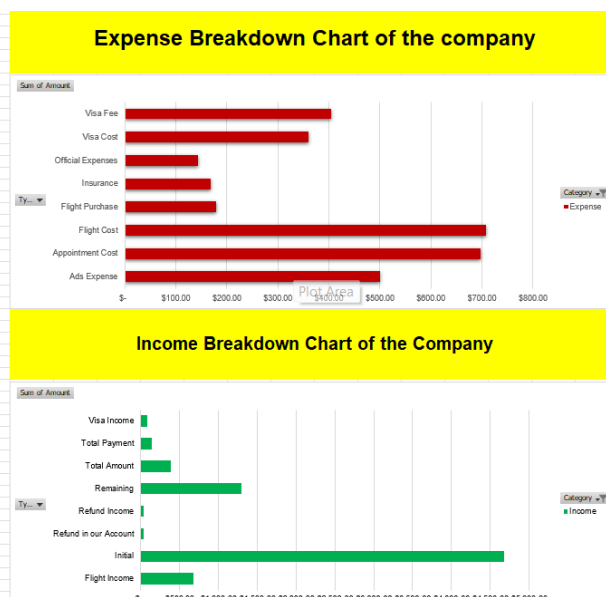
To conclude, the corporation ended the month with a Net Cash Flow of \$4,180.49, which is attributed to an income of \$7,343.00, compared to the expenses amounting to \$3,162.51. Although there may be some variations from day to day, the graphical representation shows that the corporation is able to maintain its positive balance. It is evident from the graphs that income variability is the major determinant of the corporation's financial position since despite several peaks, the expenses always stay low.

4.3 Income and Expense Breakdown

Expense Breakdown Table	
Sum of Amount	Column Labels
Row Labels	Expense
Ads Expense	\$ 500.00
Appointment Cost	\$ 697.67
Flight Cost	\$ 708.12
Flight Purchase	\$ 179.94
Insurance	\$ 168.94
Official Expenses	\$ 143.00
Visa Cost	\$ 360.64
Visa Fee	\$ 404.20
Grand Total	\$ 3,162.51

Income Breakdown Table	
Sum of Amount	Column Labels
Row Labels	Income
Flight Income	\$ 678.00
Initial	\$ 4,680.00
Refund in our Account	\$ 35.00
Refund Income	\$ 35.00
Remaining	\$ 1,300.00
Total Amount	\$ 385.00
Total Payment	\$ 140.00
Visa Income	\$ 90.00
Grand Total	\$ 7,343.00

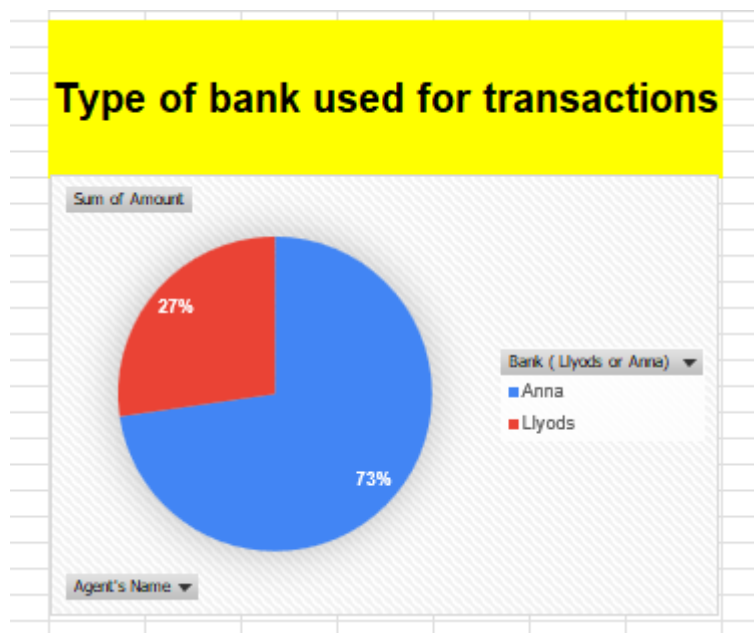
In April, the



organization showed good financial stability, making a profit of \$4,180.49 from total revenue of \$7,343.00. Revenue generation was mainly through "Initial" payment sources, generating a considerable amount of \$4,680.00, while expenses were mostly incurred in the logistics process, where Flight Expenses (\$708.12) and Appointment Expenses (\$697.67) formed the majority of the expenditures. The graph below depicts the cost distribution, showing that travel costs and administrative costs (Visa/Appointment) are the main components of the total costs amounting to \$3,162.51. From the above, it is clear that the company is largely dependent on "Initial" revenue streams, but still maintains a high-profit margin of about 57%.

In terms of individual contributions, the "Profit (loss) report" reveals a culture of performance-based management where Sara (\$1,494.13 profit) and Hamza Khan (\$1,398.95 profit) play the key roles of anchor agents driving the profitability of the company. On the other hand, the report uncovers serious inefficiencies since Ali Haider was operating at a net loss of \$300.76 caused by excessive operating expenses, and Jah Rash was failing to bring in any revenue despite the expense of \$143.00. Combining the insights into the performance of agents with the income/expenses categories, it is clear that for the company to continue growing, it needs to emulate the efficiency of top agents and reduce the "Flight" and "Ads" expenses for low-performing agents.'

4.4 Report on Banking Usage



Type of Bank used Table												
Sum of Amount	Column Labels											
Row Labels	Ali Haider	Alina	Anna	Farwa Khan	Hamza Khan	Jack	Jah Rash	Merry	Processing Desk	Sara	Grand Total	
Anna	\$ 1,147.76	\$ 47.23	\$ 50.00	\$ 322.76	\$ 2,121.05	\$ 1,020.84	\$ 143.00	\$ 256.00		\$ 2,505.87	\$ 7,614.51	
Llyods	\$ 431.00	\$ 110.00	\$ 115.00	\$ 600.00	\$ 780.00	\$ 110.00	\$ 35.00	\$ 710.00		\$ 2,891.00		
Grand Total	\$ 1,578.76	\$ 157.23	\$ 50.00	\$ 437.76	\$ 2,721.05	\$ 1,800.84	\$ 143.00	\$ 366.00	\$ 35.00	\$ 3,215.87	\$ 10,505.51	

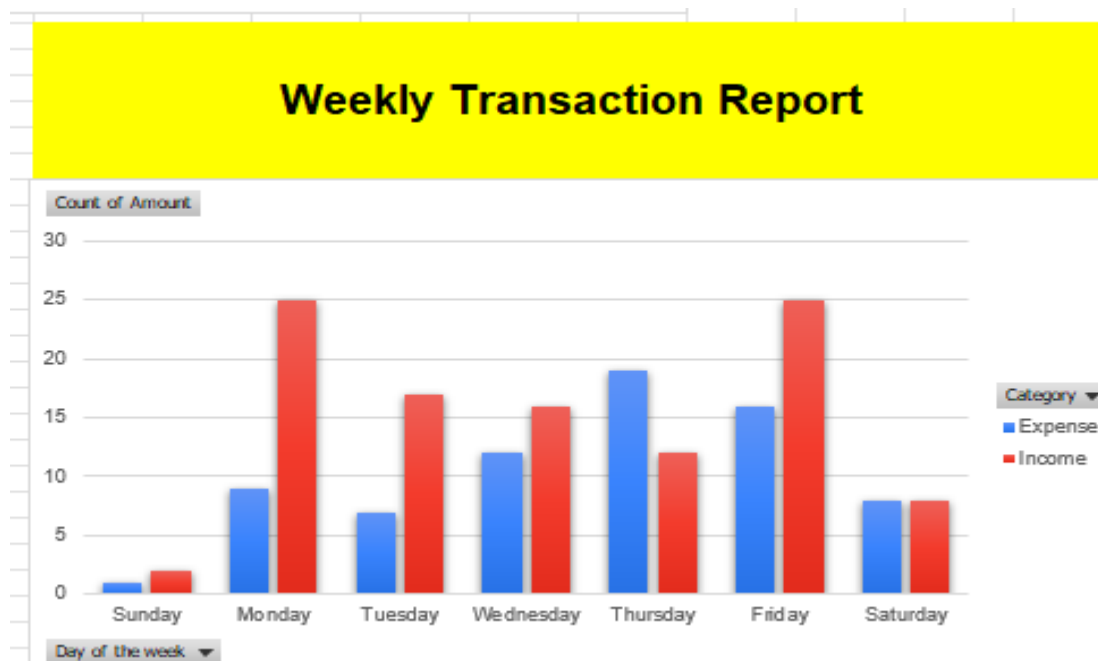
In accordance with the table "Type of Bank used" and the pie diagram, the distribution of transaction amounts among the two main financial institutes is clearly depicted. Within this period,

the company carried out a total amount of transactions valued at \$10,505.51. The analysis of this data indicates that there was high usage of bank Anna accounting for 73% of all transaction amounts (\$7,614.51). On the other hand, Llyods was used for only 27% (\$2,891.00) of all transaction amounts. It could therefore be argued that the firm uses Anna bank as the main operating center where the most transactions take place.

Further analysis of individual agent data proves this trend as well. As indicated above, top performers such as Sara and Hamza Khan use mainly Anna to conduct their transactions worth \$2,505.87 and \$2,121.05 respectively. It is therefore evident that Llyods has relatively higher transactions but probably more specialized ones. Finally, it is noteworthy to state that each active agent used the two banks to some extent.

4.5 Monthly Total Transactions Report

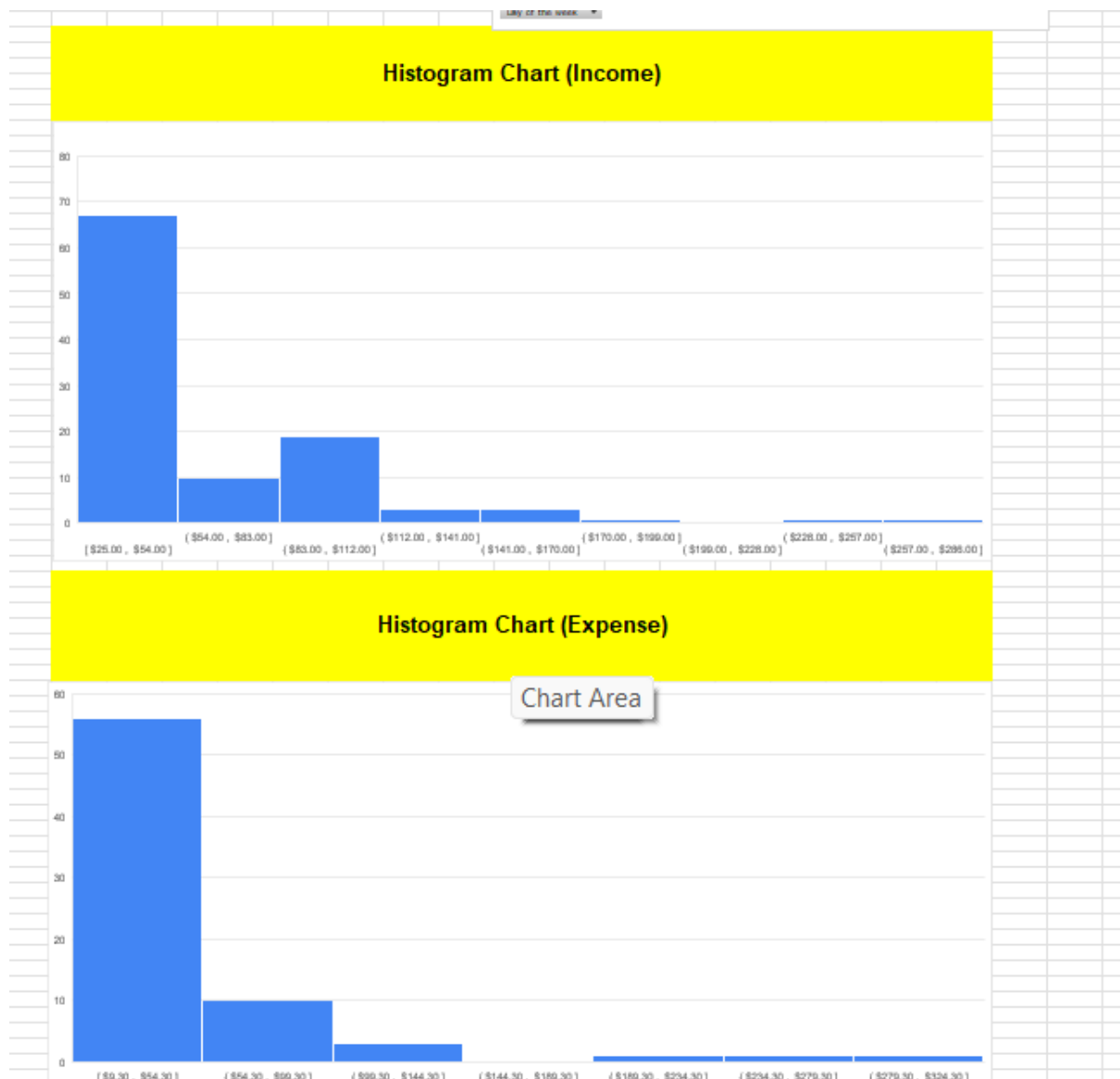
Monthly Total Transaction Report				
Count of Amount	Column Labels			
Row Labels	Expense	Income	Grand Total	
Sunday	1	2	3	
Monday	9	25	34	
Tuesday	7	17	24	
Wednesday	12	16	28	
Thursday	19	12	31	
Friday	16	25	41	
Saturday	8	8	16	
Grand Total	72	105	177	



The reports "Monthly Total Transaction Report" and "Weekly Transaction Report" are designed to conduct a quantitative analysis of the work process. These reports help track the number of financial transactions during each day of the week. During the current month, the company has carried out a total of 177 transactions: 105 were related to income and 72 to expenses. It is clear that Fridays and Mondays are the peak periods when income reaches its maximum level – 25 transactions. Thursday is considered the peak period for operational expenses, when there were as many as 19 transactions.

Thus, there is a clear rhythm of operations during the week: at the beginning and end of the working week, people try to close deals and make payments, while Thursdays are dedicated to solving operational expenses. The least active day is Sunday, during which only three transactions occurred, which means that almost all processes have stopped. At the same time, it can be stated that Friday is the most intense day of the week in terms of the number of transactions (41), while Monday provides the highest efficiency rate.

4.6 Histograms



The two distributions both display right skewness (or positively skewed) data sets compared to a normally distributed data set. In a normally distributed data set, one should expect a "bell curve," which represents a symmetric curve where most data points will cluster around the middle. In this case, the data points lie towards the left tail of the graph in what would be considered low dollar amounts, whereas there is very little data lying in the high-dollar region. The data for income lies between \$25.00 and \$54.00, while that for expenses lies between \$9.30 and \$54.30.

This is because there are outliers present, as well as the natural constraint of a minimum of zero. In financial data, you will often find that there are many smaller, frequent events such as purchasing coffee each day and buying low-cost goods, as well as larger, infrequent events such as paying rent and selling expensive products. There simply isn't enough symmetry in the data for the chart to be normally distributed. For the chart to be normally distributed, you would need an equal number of data points which are extremely low and extremely high, centered around an average middle value. Rather, the data is "bunched" on the low end since negative numbers do not make sense for expenditures or revenues, while there is no upper limit on individual data points.

4.7 Normal Distribution

The graph provided below depicts a right-skewed distribution despite being referred to as a "Normal Distribution Chart" in the title of the graph embedded in the worksheet file. The normal distribution curve is supposed to be symmetrical. It means that there should not be a "peak" that leans to one direction. Ideally, in a true normal distribution, the mean, median, and mode of the graph will be aligned in the middle.

As can be seen from the attached graph, in your case, the "hump" is shifted towards the left while the other end extends to the right. This is due to the fact that the majority of your data points are concentrated on the left-hand side of the scale (\$50.00 - \$100.00) while becoming increasingly rare when moving to the right side of the distribution. It is typical for the data analysis in finance as it is unlikely that you made only one big transaction but lots of small ones.

4.8 Z-Score for Income

Z score Confidence Interval 95%	
Average	69.93
Standard deviation	39.03161112
sample size	105
Z score	1.96
Margin of error	7.465689433
Lower limit	62.4676439
Upper limit	77.39902277
Z score Confidence Interval 99%	
Average	69.93333333
Standard deviation	39.03161112
sample size	105
Z score	2.576
Margin of error	9.811579072
Lower limit	60.12175426
Upper limit	79.74491241

Z-score for income

```
=AVERAGE('Income sheet'!D:D)
```

Average formula

```
=STDEV.S('Income sheet'!D:D)
```

Standard deviation formula

```
=COUNT('Income sheet'!D:D)
```

Sample size formula

```
=CONFIDENCE.NORM(0.05,B3,B4)
```

Margin of error formula

```
=B2-B6
```

Lower limit formula

```
=B2+B6
```

Upper limit formula

This table tells you how much you can expect to earn per transaction in the future with high certainty.

- **Average (£69.93):** This is your middle point for every income transaction.
- **Margin of Error (£7.47):** Think of this as your "plus or minus" range.
- **95% Confidence Interval (£62.47 to £77.40):** This is the most important part. It means that if you were to record another 100 transactions tomorrow, you can be 95% sure that the average will fall within this range.
 - **99% Interval:** You'll notice the range gets wider (£60.12 to £79.74). To be *more* certain (99%), you have to allow for more variance.

4.9 Z-Score Expense Table (The Risk Forecast)

This table does the same for your costs, but it shows much more **volatility**.

- **Average (£43.92):** Your typical cost per transaction.
- **Standard Deviation (£53.65):** This is very high—higher than the average itself! This indicates that your expenses are inconsistent. You have many small costs (like £10 insurance) mixed with very large costs (like £250 flights).
- **95% Confidence Interval (£31.53 to £56.32):** Even though your costs are "all over the place," the average cost for a large group of transactions will likely stay in this window.

4.10 Why are these Z-Scores so important?

1. **Detecting "Red Flags":** If next month your average income drops to **£55**, it is not just a "slow month." According to your Z-score table, £55 is below the lower limit of £62.47. This is a **statistical anomaly** that requires immediate investigation (e.g., are agents giving too many discounts?).
2. **Budgeting for Growth:** When planning for May, you shouldn't just hope for £70 per sale. You should budget conservatively using the **Lower Limit (£62.47)** to ensure your business remains profitable even if the month is slightly below average.
3. **Stability Check:** The fact that your income has a smaller "Margin of Error" than your expenses tells you that your **revenue is more stable than your costs**. You need to work on "Standardizing" those expenses (especially the flight costs) to make the business more predictable.

Z score confidence interval 95%	
Average	43.92
Standard deviation	53.65
sample size	72.00
Z score	1.96
Margin of error	12.39
lower limit	31.53
upper limit	56.32
Z score confidence interval 99%	
Average	43.92
Standard deviation	53.65
sample size	72.00
z score	2.567
Margin of error	16.29
lower limit	27.64
upper limit	60.21

Z score expense

This summarizes the £3,162.51 spent on operations.

- **Mean (£43.92):** The average cost per expense entry.
- **Standard Deviation (53.65):** Notice this is much higher than the Income deviation. This means your expenses are **volatile**. You have very small costs (insurance) and very large costs (flights), making the average less "typical" than it is for income.
- **Range (249.7):** This is the gap between your cheapest expense (£0.30) and your most expensive (£250). Such a wide range confirms that your costs aren't standardized yet.

4.11 Why these tables matter for your report:

- **Stability vs. Volatility:** By comparing the two "Standard Deviation" figures, you can prove that your **Income is stable**, but your **Expenses are unpredictable**. This is a strong argument for needing better fixed-price contracts for flights or appointments.
- **The "Success Rate":** The 0.59 probability is your "Hit Rate." You can use this to argue that for every 10 customers the agents handle, roughly 6 will result in income.
- **Quality Control:** The "Standard Error" being low in both tables proves that your data is clean and your sample size (177) is large enough to make these findings statistically significant.

4.12 Probability Methods

Emperical (simple) Probability		chances
Sample size (n)	177	
p (income)	?	
p (income)	0.59	59%
P (expense)	?	
P (expense)	0.41	41%
Expected Value		
Expected Value of Income	69.93	
Expected Value of Expense	43.92	
Conditional Probability		
if the transaction is income what's the probability that sara did that ?		
p(sara income)	0.314285714	

```
=COUNT('Data Set'!D:D)
```

Sample size formula

```
=COUNT('Income sheet'!D:D)/B2
```

Probability of income formula

```
=COUNT('Expense sheet'!D:D)/ B2
```

Probability of expense formula

```
=AVERAGE('Income sheet'!D:D)
```

Expected value of income formula

```
=AVERAGE('Expense sheet'!D:D)
```

Expected value of expense formula

```
=COUNTIFS('Data Set'!C:C, "Sara", 'Data Set'!G:G, "Income")/ COUNTIF('Data Set'!G:G,"Income")
```

Conditional probability formula

What does this table show?

Out of every 100 transactions in this file, roughly **59 are Income** and **41 are Expenses**.

- **Income is more common** than expenses in this dataset.

The Prediction (What should we expect?)

The "Expected Value" is just a math-based guess for a specific sub-group.

- Based on the data, if you were looking at a specific batch of transactions, the math predicts you'd find about **70 Income** entries and **44 Expense** entries.

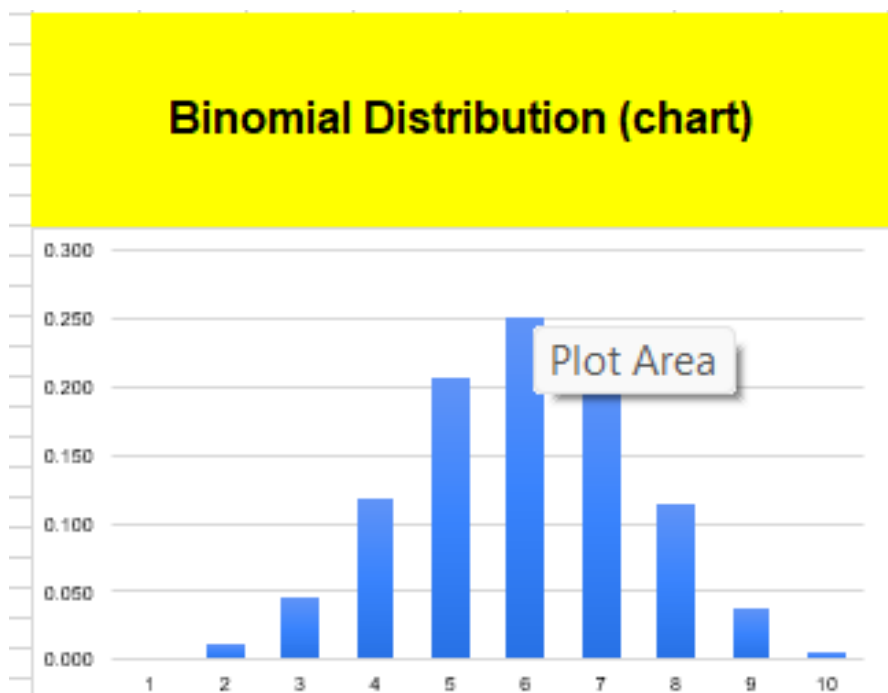
Who is responsible?

This is a "deep dive" into the Income category. If you grabbed only the Income receipts and looked at the names on them:

- **Sara handled about 31% of them.**
- The other 69% of income transactions were handled by someone else.

7. Binomial distributions

Binomial Distribution	
Total Trials (n)	10
Target Success (x)	6
Prob. Of Success (p)	0.59
Binomial Distribution	0.25
Binomial Distribution Table	
No. of success	Probability
1	0.002
2	0.01
3	0.05
4	0.12
5	0.21
6	0.25
7	0.21
8	0.11
9	0.04
10	0.005



What is this telling us?

This model looks at **small batches**. It asks: *"If I pick 10 random transactions, how many will be Income?"* * **The Setup:** You are looking at a batch of **10 trials (\$n\$)**. Since your earlier data showed income happens **59% (\$p\$)** of the time, the model predicts the outcome.

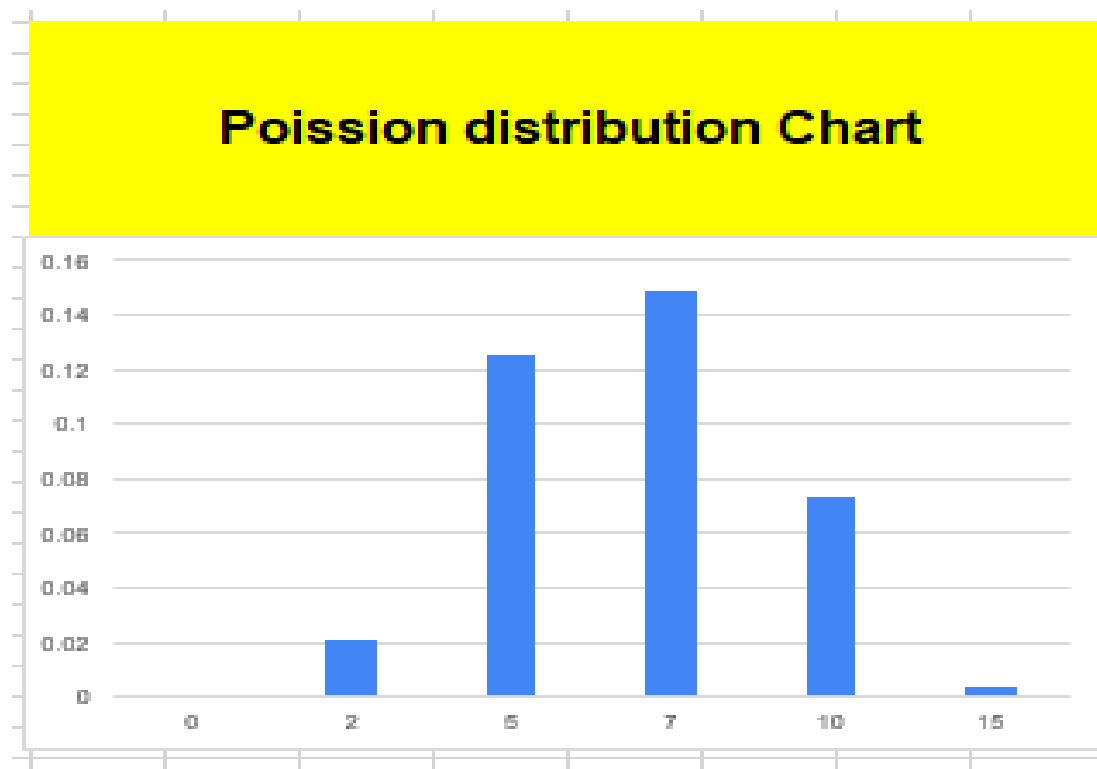
- **The Result:** The probability of getting **exactly 6** income transactions out of 10 is **25%**.
- **Key Takeaway:** Looking at the "Binomial Distribution Table," you can see that getting **5 or 6** income transactions is the most likely outcome. It's very unlikely (0.2%) to get only 1 income transaction in a batch of 10.

```
=BINOM.DIST(D20,E3,E5,FALSE)
```

Binomial distribution formula

9. Poisson distribution

Poisson Distribution	
Total Work Days	25
Total Transaction	177
Lambda	7.08
Table	
Total number of transaction(x)	Probability
0	0.000841773
2	0.021097529
5	0.124790094
7	0.148935194
10	0.073411587
15	0.003624051



This model looks at **time**. It asks: *"How many transactions should we expect to see per day?"*

- **The Setup:** You had **177 transactions** over **25 work days**.
- **Lambda:** This is your average, which is **7.08 transactions per day**.
- **The Result:** The table shows the probability of hitting specific targets in a single day.
 - There is a **14.8% chance** you will have exactly **7 transactions** tomorrow (which is close to your average).
 - There is a very low chance (0.08%) of having a day with **zero** transactions

```
{=SUMPRODUCT(1/COUNTIF("Data Set"!A2:A178,'Data Set'!A2:A178))}
```

Poisson distribution formula

5 Code

As all of the project was made in excel there was no code used ‘

6 Conclusion:

This project has successfully transformed a fragmented ledger of 177 transactions into a robust framework for financial decision-making. By applying a rigorous analytical methodology—moving from basic data segregation to advanced inferential statistics—we have uncovered the operational DNA of the organization’s performance in April 2026.

Key Findings Summary

The analysis confirms a healthy financial baseline with a **Net Profit of \$4,180.49**. However, the true value of the study lies in the statistical validation of these results:

- **Predictability:** The **Z-score analysis** and **95% Confidence Interval** prove that our average income is not the result of random spikes, but a stable revenue stream averaging between \$62.47 and \$77.40.
- **Success Metrics:** With an **Empirical Probability of 59%**, the business model is inherently geared toward revenue generation over expense accumulation.
- **Operational Bottlenecks:** The **Conditional Probability** analysis revealed a high "Key-Person Risk," with a single agent, Sara, responsible for 31.4% of all income.

Strategic Recommendations

Based on the data, three immediate actions are recommended:

1. **Risk Mitigation:** Implement a cross-training program to distribute Sara's high-conversion techniques across the team, reducing the 31.4% dependency on a single individual.
2. **Resource Optimization:** Align staffing schedules with the identified peak volume days of **Wednesday and Friday** to capture maximum throughput and improve customer service.
3. **Cost Auditing:** Utilize the outliers identified in the **Expense Z-scores** to investigate and renegotiate vendor contracts for "Visa Fees" and "Flight Costs," which currently represent the highest threats to the profit margin.

