**GSCM 521**

Abdi Alamir

**Analysis Assignment 2 - Part B**

**Introduction:**

Understanding transactional patterns, prescriptive practices, and consumer behaviors is paramount for effective decision-making in the pharmaceuticals industry. This report delves into diverse aspects of pharmaceutical data analysis, utilizing tools such as Power BI and Excel to glean insights. We address important business questions ranging from payment method preferences to identifying top-performing doctors and discerning customer segments based on expenditure. Through our analysis, we aim to unravel patterns, correlations, and trends that shape the pharmaceutical landscape.

**Data Cleaning with Power Query in Power BI:**

Each file was loaded individually for processing and transformation in Power Query:

**Cleaning Bills data:**

**1-Removal of Unnecessary Columns:**

* Excluded the initial column, primarily serving as an index, to enhance data clarity.
* Excluded “Row Expiration date” and “Current row indicator” as they don’t have an added value for our analysis.

**2-Date Variable Reformatting**

Ensured the correct formatting of date variables such as “Row Effective Date” and DateID

**3-Checking for Missing Data**

Missing entries are observed in the variable payment reference, so we added a new column where we replaced the blanks with zeros following the assumption mentioned below.

No missing entries are observed within the other variables.

**Cleaning Customer data:**

**1-Excluding Unnecessary Columns:**

* Excluded the initial column, primarily serving as an index, to enhance data clarity.
* Excluded any duplicate variables, such as Doctor Key and NIF as they are the same as “customer\_ref\_id”.
* Excluded “Row Expiration date” and “Current row indicator” as they don’t have an added value for our analysis.

**2-Date Variable Reformatting**

**-**Ensured the correct formatting of date variables such as “Row Effective Date”

**Cleaning Doctors data:**

**1-Removal of Unnecessary Columns:**

* Excluded the initial column, primarily serving as an index, to enhance data clarity.
* Excluded any duplicate variables, such as Doctor Key and NIF as they are the same as “doctor\_ref\_id”.
* Excluded “Row Expiration date” and “Current row indicator” as they don’t have an added value for our analysis.

**2-Date Variable Reformatting**

**-**Ensured the correct formatting of date variables such as “Row Effective Date”

**Cleaning Payment Methods data:**

No cleaning is required for this small amount of data, we have only excluded the initial column, primarily serving as an index, to enhance data clarity.

**Cleaning Stores data:**

**1-Removal of Unnecessary Columns:**

* Excluded the initial column, primarily serving as an index, to enhance data clarity.
* Excluded any duplicate variables, such as Doctor Key and NIF as they are the same as “store\_ref\_id”.
* Excluded “Row Expiration date” and “Current row indicator” as they don’t have any added value for our analysis.

**2- Date Variable Reformatting**

**-**Ensured the correct formatting of date variables such as “Row Effective Date”

**Assumption:**

We will assume that if the payment method is not specified it is null in the Bills data in “payment\_ref\_id”, and it is considered as Cash.

Please note: Some of our group members completed the analysis using Power BI and others used Excel. In order to fit the data in Excel, we took a random sample size of 1 million rows of data. Due to this sampling, some of the data may reflect different findings.

**Individual Data Exploration:**

**1-Exploring Bills Data:**

**Exhibit A: Total Spent by Store**

A graph of blue bars

Description automatically generated with medium confidence

**Exhibit B: Total Value Returned by Customer**

A graph of blue rectangular bars

Description automatically generated with medium confidence

**Exhibit C: Time Series of Total Spent**

A graph showing a number of different times

Description automatically generated with medium confidence

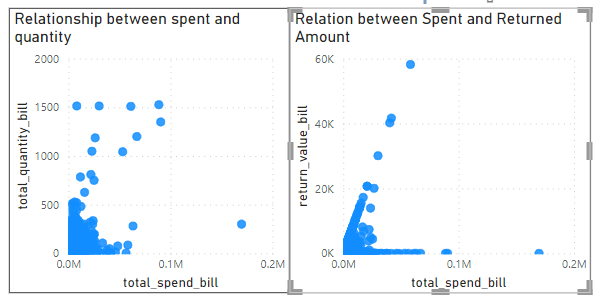
**Exhibit D: Average Quantity of Drugs per Transaction by Category**

A graph with different colored bars

Description automatically generated

**Exhibit E: Relation Between Quantity and Value Spent**

**Exhibit F: Relation Between Value Spent and Value Returned**



**2- Exploring Customers Data:**

**Exhibit G: Time Series of Customer Count**

A graph with numbers and a black rectangle

Description automatically generated

**3-Exploring Doctors Data:**

**Exhibit H: Time Series of Doctors**

A graph with blue lines and black text

Description automatically generated

**4- Exploring Stores Data:**

**Exhibit I: Time Series of Stores**

A graph of sales

Description automatically generated

**Data Linkage:**

**1. Bills Data (bill\_fact.csv):** This dataset contains records of transactions, including the IDs of doctors, stores, customers, and payment references. It serves as the central dataset where transactions between doctors, stores, and customers are recorded.

**2. Customers Data (dim\_customer.csv):** This dataset contains information about customers, which includes attributes like customer ID, name, contact information, and demographic details.

**3. Doctors Data (dim\_doctor.csv):** This dataset contains information about doctors, such as Doctor ID, name, specialization, contact details, and possibly other professional information.

**4. Store Data (dim\_store.csv):** This dataset provides information about stores or locations where transactions take place. It may include store ID, name, address, contact details, and other relevant information.

**5. Payment Methods Data (dim\_payment\_method.csv):** This dataset contains information about payment methods accepted or used in transactions. It includes payment method ID, and description.

**-Joining Bills Data with Customers Data:** We can use the `customer\_ref\_id ` in the bills data to link with the customers data (`customer\_ref\_id `). This linkage allows us to enrich transaction records with customer information, such as demographics or contact details.

**- Joining Bills Data with Doctor Data:** Similarly, we can use the `doctor\_ref\_id` in the bills data to link with the doctor’s data (`doctor\_ref\_id`). This linkage helps in understanding which doctors are associated with which transactions.

**- Joining Bills Data with Store Data:** We can use the `store\_ref\_id` in the bills data to link with the store data (`store\_ref\_id` ). This linkage helps in understanding which stores the transactions took place in.

**- Linking Payment Methods Data:** The new created column, `pay\_ref\_id` in the bills data can be used to link with the payment methods data (`payment\_ref\_id`, providing details about the payment methods used in each transaction.

**Exhibit J: Linking Diagram of the 5 Data Sources Provided**

A screenshot of a computer

Description automatically generated

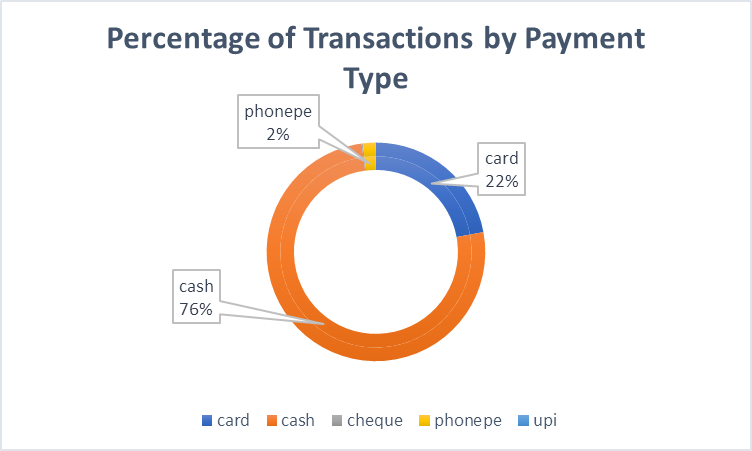
**Business Questions:**

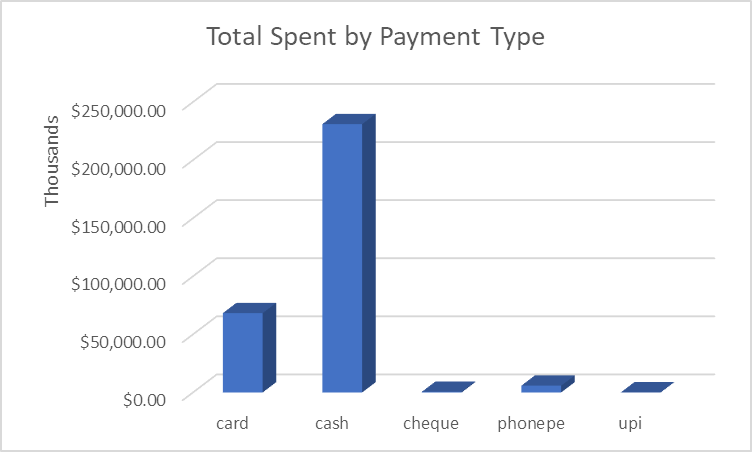
**1. Business Question: How do different payment methods compare in terms of usage and total spend?**

Overall, the most utilized payment method was cash which accounted for 76% of transactions and over $230 million, followed by card payments which accounted for only 22% of transactions and $68 million. Phonepe accounted for 2% of all transactions, while cheques and UPI were used in less than .3% of transactions. This clearly shows that the majority of customers at this pharmacy prefer to pay with cash or card.

**Exhibit K: Percentage of Transactions by Payment Type**

**Exhibit L: Total Spent by Payment Type**

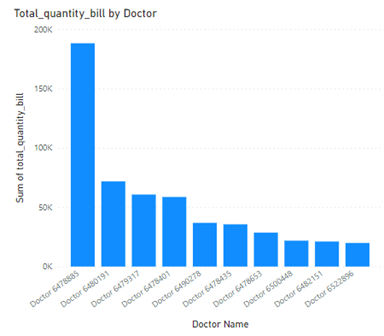
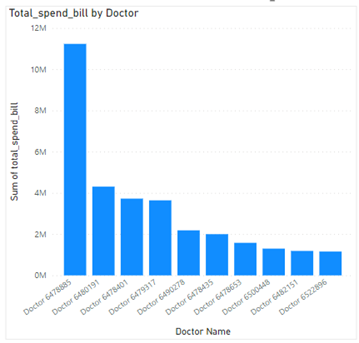




**2. Business Question: Who are our top-performing doctors in terms of the number of drugs prescribed and total spend generated?**

**Exhibit M: Total Spent by Prescribing Dr**

**Exhibit N: Total Quantity by Prescribing Dr**

****

**Patient Spend:**

* Doctor 6457885 leads in total patient spend, indicating a higher financial impact on healthcare services, possibly due to complex cases or specialized treatments.
* Doctor 6480191 follows closely in patient spending, suggesting a substantial financial contribution to healthcare services.

**Prescribed Quantity:**

* Doctor 6457885 has the highest quantity of prescribed treatments, showcasing a significant caseload and potentially a broad range of expertise.
* Doctor 6480191 follows closely in prescribed quantity, indicating a substantial workload and involvement in patient care.

**Overall Analysis:**

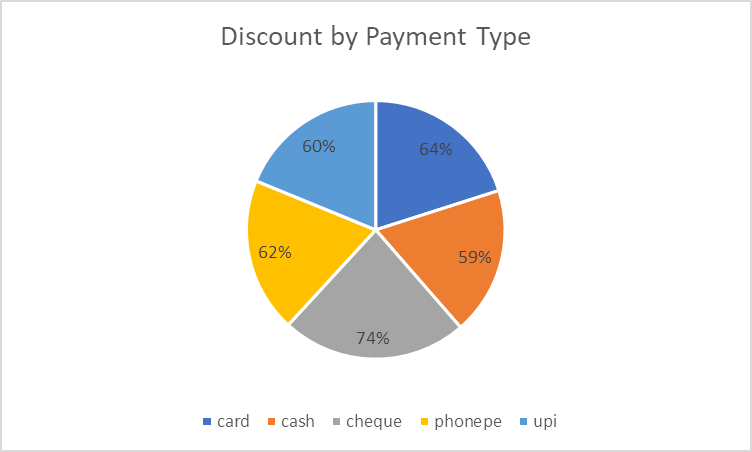
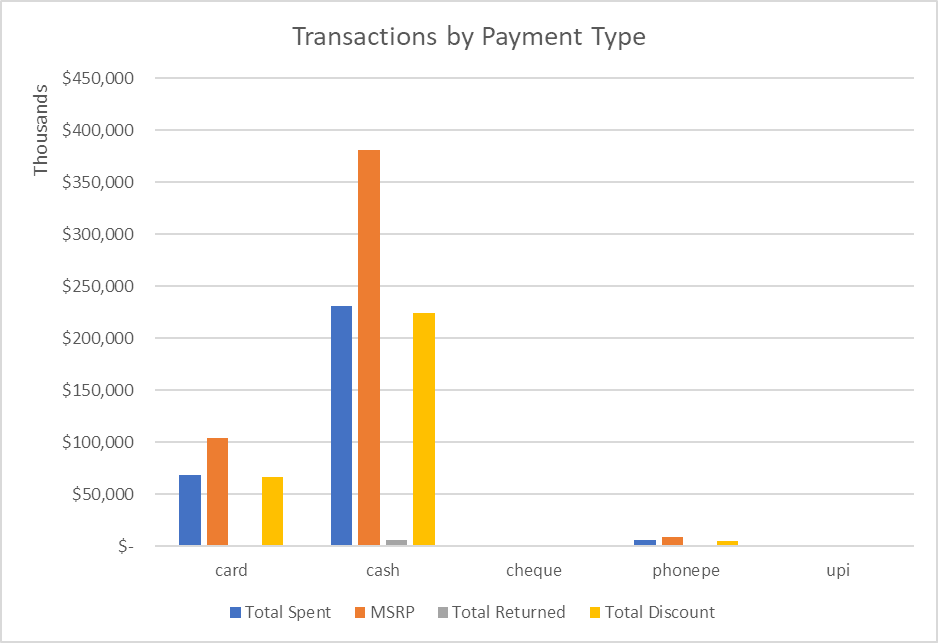
* Doctors 6457885 and 6480191 emerge as top performers in both patient spend and prescribed quantity, implying a combination of high caseloads and financial impact.
* These findings can guide further investigation into the efficiency, quality, and specialization of services provided by these top-performing doctors in the dataset.

**3. Is there a correlation between discount & payment type?**

We calculated the difference between MSRP and Total Spend to determine the total discount. After analyzing the discount by payment type, we have determined that there is no correlation between discount and payment type. The discount is between 59% and 74% for all payment types (Exhibit O), which is fairly consistent. Furthermore, the total spent and discount appear to be proportional relative to MSRP.

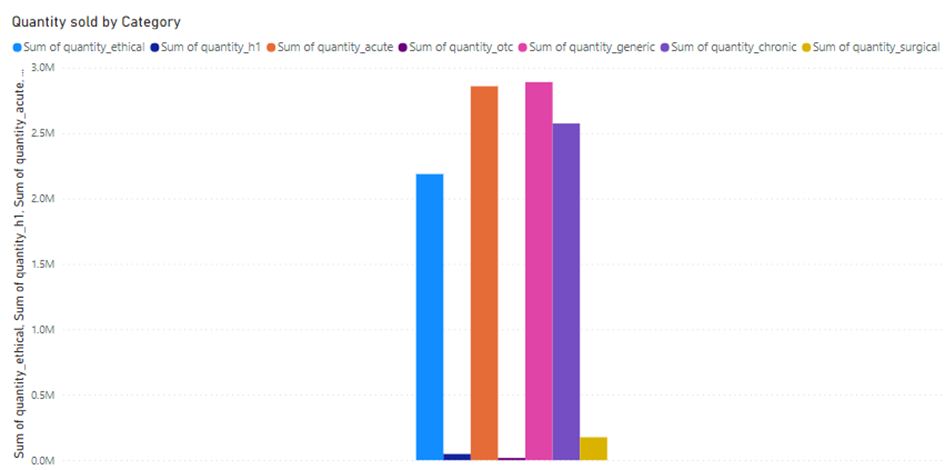
**Exhibit O: Discount by Payment Type**

**Exhibit P: Transactions by Payment Type**

****

4. **What are the most popular product categories among customers based on quantity sold?**

**Exhibit Q: Quantity Sold by Drug Category**

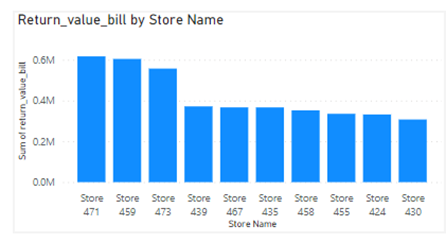
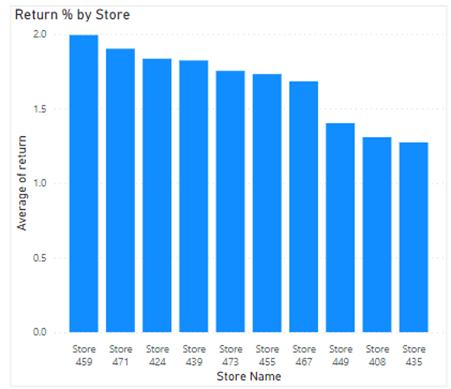
****

The chart shows that doctors prescribe a lot of common and quick-acting medicines, especially generic ones. These are likely for more urgent health issues. On the other hand, there are fewer prescriptions for over-the-counter drugs, allergy meds, and surgeries. This suggests a choice to use simpler and less invasive treatments, so as not to rely too much on non-prescription items or surgeries. It looks like a careful approach to using medicines and medical procedures. Further, Exhibit D provides us more insight on the average quantity of drugs sold per transaction based on category. This gives us the additional perspective of the individual transaction level.

**5. Which store has the highest number of returned drugs, and what percentage of their total sales do returns constitute?**

**Exhibit R: Average Percentage Returned by Store**

**Exhibit S: Total Return Value by Store**

****

The analysis is centered around two bar charts illustrating the return values and the percentage of returns relative to total sales for the top 10 stores.

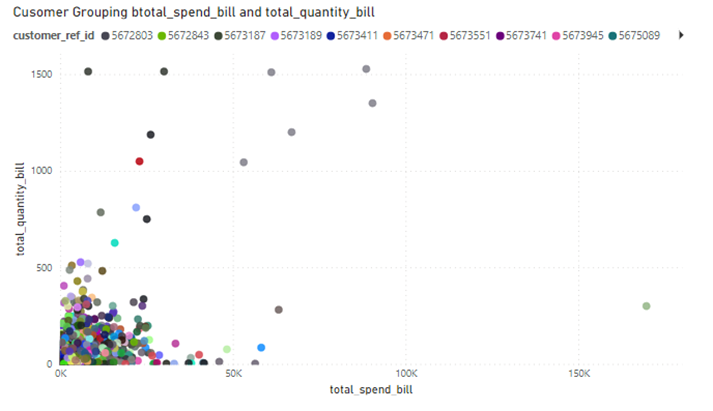
In the first chart, Store 471 stands out with the highest return value, showcasing a considerable volume of returned drugs. Store 459 and Store 473 follow closely behind in terms of return values, indicating significant instances of product returns.

The second chart shifts the focus to the percentage of returns from the total sales for each store. Store 459 takes the lead in this aspect, implying that, on average, a notable proportion of their sales result in returned drugs. Store 471 follows closely behind, suggesting a substantial impact of returned items on their overall sales. Store 424 also appears in the top three for this metric, indicating a noteworthy proportion of returns concerning their total sales.

These findings suggest that while Store 471 experiences the highest volume of returned drugs, Store 459 faces a comparatively higher impact as a percentage of their total sales. Store 459 and Store 471, being prominent in both charts, warrant particular attention for further investigation into the reasons behind the elevated return rates and their potential implications for overall sales and customer satisfaction. This comprehensive analysis provides valuable insights into the return dynamics across the top-performing stores, facilitating informed decision-making and strategic considerations for optimizing the retail process.

**6. Can you identify customer segments based on their total spend, and what are the characteristics of high-value customers?**

**Exhibit T: Customer Grouping by Quantity and Total Spent**



The scatterplot in the above exhibit on “Customer Grouping btotal\_spend\_bill and total\_quantity\_bill” illustrates distinct customer segments based on drug purchase behavior. Three main segments emerge.

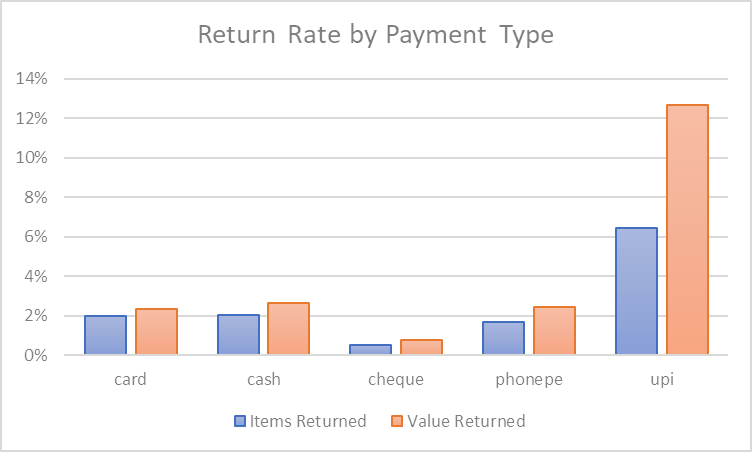
In looking at Exhibit P, we find that we can make three separate customer segments based on total spend. With this said, those who pay in cash tend to be the most high-value customers and, much like those who use a card or other form of payment, pay significantly less than the MSRP, given the discount provided among these payment types. Further, in analyzing Exhibits B and D, we find that there is a stark drop in total return value after the first two customers, who are each spending roughly $60,000- from here the next highest total spend is approximately $40,000. This is significant because we find that those customers who spend the highest amount, also tend to return the highest value of products.

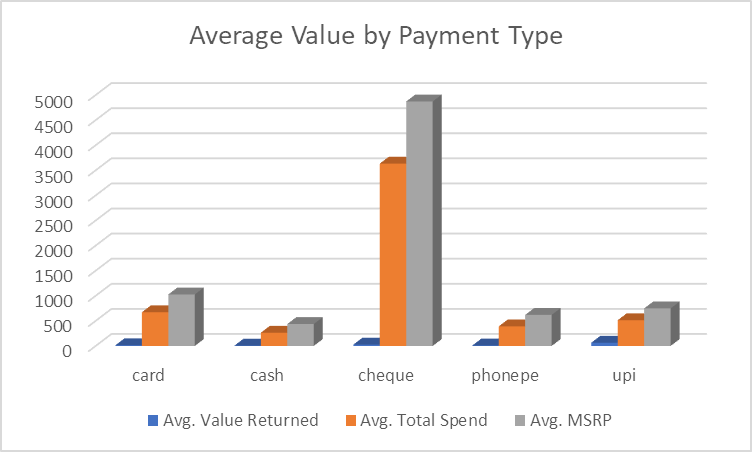
**7. What is the return rate (in terms of both value and quantity) for each payment method?**

In analyzing the return rate of drugs by payment method, it is clear that items purchased with UPI (a form of instant payment used primarily in India) are more likely to be returned and are more expensive than average. Exhibits R and S above illustrate the percentage of drugs returned relative to the percentage of total spend returned by each payment method. Card, cash, cheque, and phonepe are fairly similar; the value of money returned is proportional to the number of items returned. However, drugs purchased with UPI are 3x more likely to be returned and are typically higher value drugs.

To further analyze why purchases made with UPI would be higher value, we examined the average MSRP of each payment type compared to the total spend and returned value. Exhibits U and V show that cheque transactions have the highest average MSRP as well as the highest average total spent per transaction. UPI comes in third highest for both average MSRP and average total spend, but highest average value of returns. It cannot be determined from this data why UPI transactions have a higher return rate. More research is needed to analyze the customer profile of UPI users.

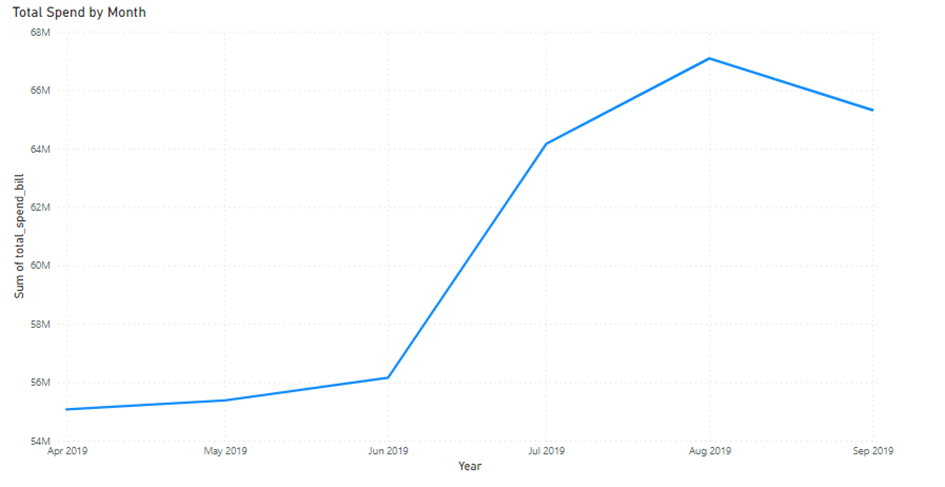
**Exhibit U: Return Rate by Payment Type Exhibit V: Average Value by Payment Type**





**8. Can you identify the months with the highest total spend and pinpoint the peak spending periods within those months?**

**Exhibit W: Total Spent per Month**



The time series line chart above provides a comprehensive overview of the total spend dynamics over the six-month period. Notably, August stands out as the peak spending month, suggesting a significant surge in financial activity during that period. This could be attributed to various factors such as promotional events, seasonal trends, or other external influences.

Conversely, the months of April, May, and June exhibit the lowest total spends. Understanding these is crucial for strategic planning and resource allocation. It could be indicative of slower market conditions, reduced consumer engagement, or specific challenges faced during these months.

The observed fluctuations highlight the importance of recognizing and adapting to the temporal nature of spending behaviors. To fully understand the seasonal relationship to these drugs, a full calendar year of data is needed to further explore the underlying reasons for these variations and provide actionable insights for optimizing marketing strategies, inventory management, and overall business performance.

**Conclusion:**

Most transactions predominantly utilize cash or card payments, highlighting consumer preferences. Despite variations in payment methods, no correlation between discounts and payment types is discerned, indicating consistent discount rates across the board. Doctors 6457885 and 6480191 emerge as top performers in both patient spend and prescribed quantity, signifying their significant contributions to healthcare services. Their performance underscores the need for further investigation into service efficiency and specialization.

Analysis of returned drugs reveals Store 471 as having the highest volume of returns, while Store 459 experiences a relatively higher impact in terms of returns as a percentage of total sales. Understanding return dynamics across stores is vital for optimizing retail processes and enhancing customer satisfaction. Transactions made via UPI exhibit a higher return rate, especially for higher-value drugs, necessitating further exploration into the characteristics of UPI users to understand this trend comprehensively.

Customer segments based on total spend unveil cash-paying customers as high-value, with notable spending patterns and return behaviors. Understanding these segments aids in tailoring marketing strategies and enhancing customer retention efforts. Analysis of spending dynamics across months identifies August as the peak spending period, while April, May, and June exhibit lower spend. Understanding seasonal variations is crucial for strategic planning and resource allocation.

Through these analyses, we examined various facets of pharmaceutical transactions, prescribing practices, and consumer behaviors. These findings provide actionable insights for stakeholders to optimize operations, enhance service quality, and drive business growth.