Reducing Food Wastage and Procurement Loss in the IISER Thiruvananthapuram Student Mess Using Data Warehousing and Business Intelligence Techniques: A Varsha 2025 Case Study

by

Maitreya Sameer Ganu
IMS23099
Course Project, January–April 2026
Indian Institute of Science Education and Research,
Thiruvananthapuram

Supervisor: **Dr. Zakaria Laskar**
School of Data Science, IISER Thiruvananthapuram

January 28, 2026

Contents

Chapter 1

Introduction and Data Availability

Food wastage is a significant operational and sustainability challenge in institutional dining facilities such as student messes. Despite fixed consumption patterns, large quantities of cooked food are wasted daily due to over-procurement, poor demand estimation, and lack of data-driven decision-making.

This project aims to reduce food wastage in a student mess by applying Data Warehousing and Business Intelligence (DWBI) techniques to real-world mess expense data. Since item-level wastage data was unavailable, expense data was used as a proxy to identify inefficiencies and wastage-prone periods. This case study examines the IISER Thiruvananthapuram mess, a student-run, non-profit food service system, with the objective of reducing food waste and the associated financial losses. Given the non-profit nature of the mess committee, the primary focus of this study is not profit maximization but operational efficiency—specifically minimizing avoidable food waste while maintaining food quality and meeting student demand. By identifying inefficiencies in procurement, consumption patterns, and demand forecasting, the study aims to propose data-driven strategies that enable cost savings through waste reduction rather than revenue expansion. Since direct food wastage measurements were unavailable, a proxy-based estimation approach was adopted. A baseline wastage of 100 kg per day was assumed, and estimated wastage was computed by scaling this value according to the ratio of daily expense to average daily expense. This method identifies the relative risk of waste rather than the absolute amounts of waste. It is important to note that the daily expense values shown in Table 5 do not correspond to single transactions. The raw dataset records expenditures at the vendor-bill level, meaning that multiple vendor payments may occur on the same calendar date. Daily expense values were therefore computed by aggregating all vendor-level transactions corresponding to the same date using SQL group-by operations. Raw data can be found here - here

Chapter 2

Data Warehousing, Preprocessing, and Knowledge Discovery Pipeline

This project follows a complete data warehousing and knowledge discovery workflow, transforming raw transactional mess expenditure records into analytically meaningful insights. The overall methodology aligns closely with the classical data mining pipeline consisting of data cleaning, integration, selection, transformation, mining, pattern evaluation, and knowledge presentation.

The initial stage of the workflow involved data cleaning, where raw expenditure records extracted from the IISER Thiruvananthapuram student cooperative mess portal contained inconsistencies such as missing payment dates, non-standard bill date formats, and textual vendor identifiers. These inconsistencies were resolved through systematic preprocessing using SQL, including date correction, normalization of vendor names, and removal of structurally invalid entries. This step ensured that downstream analytical operations were performed on reliable and consistent data.

Following data cleaning, data integration was performed by consolidating multiple monthly datasets (September to December 2025) into a unified relational schema. Although the source data originated from a single operational system, integration was required across time periods and logical entities such as vendors, mess units, and dates. This integration was achieved through a dimensional data warehouse design, consisting of fact and dimension tables, enabling consistent analysis across months.

Data selection was carried out using structured SQL queries to extract only analytically relevant attributes, such as payment date, vendor, mess unit, and transaction amount. These queries filtered the operational data into a subject-oriented format suitable for analysis, ensuring that irrelevant transactional metadata did not interfere with analytical clarity.

The transformation phase involved aggregating transactional-level data into daily totals, mess-group averages, and vendor-level summaries. Derived attributes such as estimated food wastage were computed using domain-informed proportional assumptions linking expenditure magnitude to wastage volume. These transformations converted raw transactional data into higher-level analytical indicators suitable for trend detection and comparative analysis.

Data mining and pattern extraction were performed primarily through temporal trend analysis and statistical distribution analysis. Monthly expenditure patterns revealed strong correlations with academic calendar events such as mid-semester and end-semester examinations, science festivals, and campus-wide events. Additionally, Benford's Law was applied at the vendor level to evaluate the naturalness of first-digit distributions in expenditure amounts, serving as a preliminary fraud detection heuristic within a business intelligence context.

Pattern evaluation was conducted by comparing observed trends with known institutional schedules and expected consumption behaviors. Peaks in expenditure and wastage coincided with increased campus occupancy, validating the plausibility of detected patterns. Deviations from expected statistical distributions were not interpreted as definitive fraud indicators but were flagged as candidates for deeper administrative review.

Finally, knowledge presentation was achieved through structured tabular summaries, time-series plots, bar charts, and comparative visualizations. These visual representations transformed complex numerical patterns into interpretable insights, enabling stakeholders to understand consumption dynamics, wastage behavior, and expenditure anomalies. Collectively, this pipeline demonstrates the practical application of data warehousing and data mining principles to a real-world institutional food systems problem.

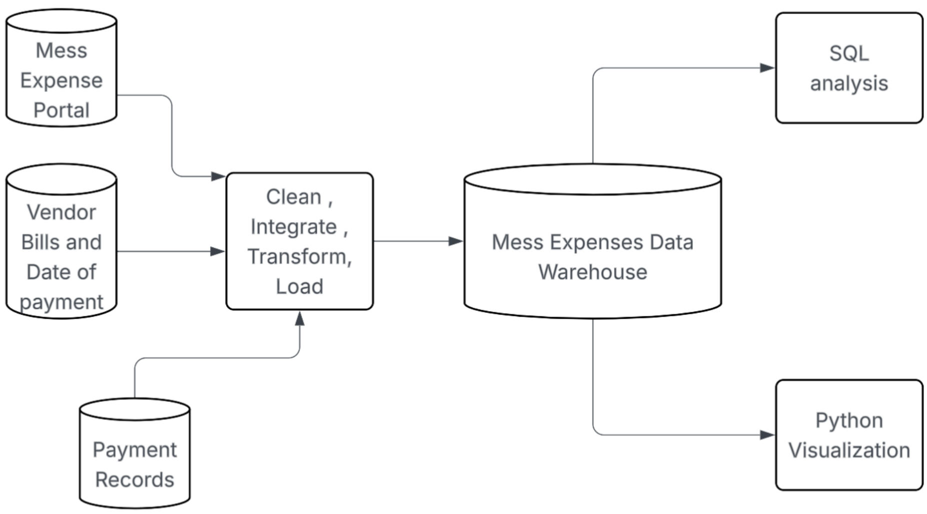


Figure 2.1: OLTP diagram of the database

Chapter 3

Design and Structure of the Mess Expense Data Warehouse

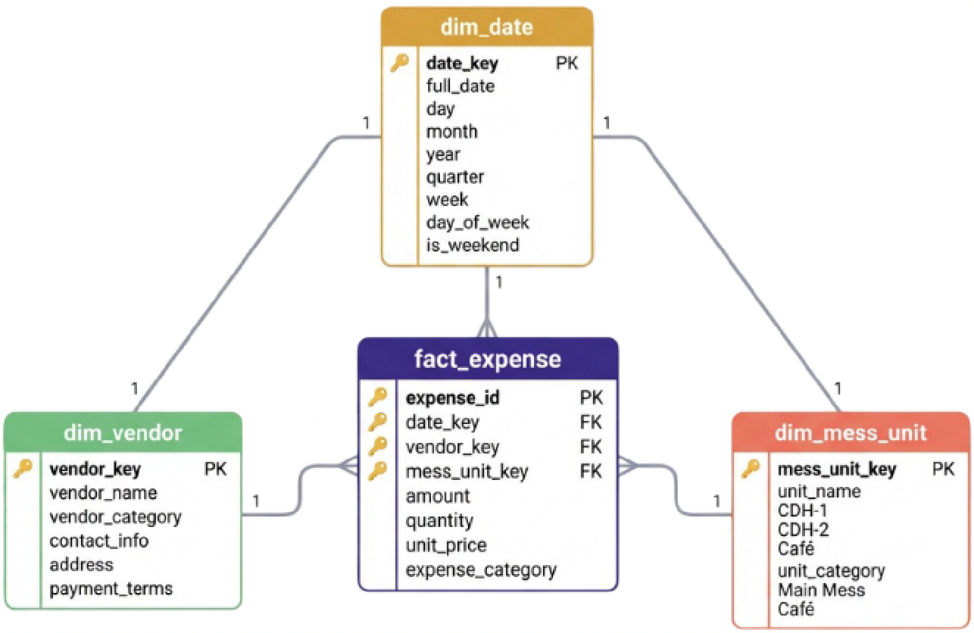


Figure 3.1: Star schema diagram showing fact table and dimensions tables

Table Name	**Table Type**	**Primary / Foreign Keys**	**Description and Key Attributes**
`fact_expense`	Fact Table	**PK:** expense_id **FKs:** date_id, vendor_id, mess_unit_id	Stores transactional expense records. Each row represents a payment made to a vendor by a specific mess unit on a given date. *Measures:* total_amount
`dim_date`	Dimension	**PK:** date_id	Provides temporal context for expense records. *Attributes:* full_date, day, month, year, day_name, is_weekend
`dim_vendor`	Dimension	**PK:** vendor_id	Stores standardized vendor information used for procurement. *Attributes:* vendor_name
`dim_mess_unit`	Dimension	**PK:** mess_unit_id	Represents operational dining units. *Attributes:* mess_name (CDH-1, CDH-2, Café), mess_category (Main Mess / Café)
`stg_expense_raw`	Staging	–	Temporary storage for raw monthly CSV uploads. Mirrors source structure and is used for cleaning, validation, and transformation prior to warehouse loading.

Table 3.1: Table 3.1 summarizes the structure of the mess expense data warehouse. Each table is described in terms of its role in the star schema, primary keys, foreign key relationships, and core attributes.

Chapter 4

Limitations and Data Quality Considerations

While the data warehouse and analytical framework developed in this study enable comprehensive analysis of mess expenditure patterns, it is important to explicitly acknowledge certain limitations and data quality considerations inherent to the available data. Addressing these aspects strengthens the credibility of the analysis and ensures that interpretations are made within an appropriate methodological context.

Payment Date versus Consumption Date Mismatch

The primary limitation arises from the fact that the available transactional data records *payment dates* rather than actual *food consumption dates*. In the operational workflow of the student mess, vendor payments are often settled in batches, with multiple bills corresponding to several days or weeks of consumption being cleared on a single payment date. Consequently, spikes observed in daily expenditure values do not necessarily indicate unusually high food consumption on those specific dates, but rather reflect the settlement of accumulated liabilities.

All temporal analyses in this study therefore interpret expenditure patterns as financial settlement behavior rather than direct consumption behavior. This distinction is consistently maintained throughout the analysis to avoid misrepresentation of daily food demand.

Aggregation Bias in Daily Expenditure Analysis

The aggregation of multiple vendor payments into a single daily total introduces aggregation bias, particularly when comparing daily expenditure values across the semester. Days with exceptionally high totals may represent delayed or bulk payments, while days with low or zero expenditure may still correspond to normal levels of food consumption that were not immediately settled financially.

Despite this limitation, aggregation remains appropriate for business intelligence purposes, as the objective of the study is to analyze financial flows, budgeting behavior, and expenditure risk patterns rather than precise per-day consumption quantities.

Proxy-Based Wastage Estimation Assumptions

Food wastage in this study is estimated using a proxy model based on total expenditure values, due to the absence of direct wastage measurement data. This approach assumes a proportional relationship between expenditure and potential wastage, enabling relative comparison across dates, mess units, and academic periods.

It is important to note that the estimated wastage values should be interpreted as *risk indicators* rather than absolute measurements. The model is designed to identify periods of unusually high financial outflow that may warrant closer operational scrutiny, rather than to quantify exact food waste in kilograms.

Lack of Item-Level Transactional Data

The dataset does not contain item-level details such as quantities of specific food items procured, menu composition, or per-meal consumption statistics. As a result, the analysis operates at the vendor and mess-unit level, limiting the ability to attribute expenditure changes to specific ingredients or menu decisions.

While item-level data would enable more granular operational optimization, the current vendor-level aggregation remains sufficient for strategic financial analysis, vendor dependency assessment, and structural network analysis of procurement relationships.

Overall Impact on Analytical Validity

Despite these limitations, the data quality is adequate for the stated objectives of the project, which focus on financial behavior, expenditure trends, structural dependencies, and anomaly detection rather than precise consumption accounting. The consistent application of assumptions across all months and mess units ensures internal comparability, making the derived insights reliable within the defined analytical scope.

Explicit recognition of these limitations enhances the transparency and robustness of the study and provides a clear foundation for future extensions involving higher-resolution operational data.

Chapter 5

August 2025 - The Semester Begins

Food wastage is a persistent operational challenge in large institutional mess facilities due to uncertainty in daily demand, over-procurement, and variability in consumption patterns. At IISER Thiruvananthapuram, informal operational observations suggest that approximately 100 kg of food is wasted per day. However, direct measurement of item-wise food wastage is not systematically recorded. This chapter applies a data warehousing and business intelligence (DWBI) framework to analyze mess expenditure data for August 2025 and to use expense as a proxy indicator for food wastage.

The raw dataset was obtained from the IISER Thiruvananthapuram Student Cooperative Mess expense records and consists of 276 vendor payment transactions for August 2025. Each record contains the vendor name, bill information, and the total amount paid. Since the dataset lacked explicit mess unit identifiers and a dedicated date column, preprocessing was required before analysis. The data was ingested into a MySQL staging table, where expense dates were extracted from bill information using SQL string operations and converted into a standard `DATE` format. Mess units were inferred using rule-based classification on vendor names, mapping records to CDH-1, CDH-2, or Café. Transactions that could not be reliably classified, such as cash vouchers and shared administrative expenses, were labeled as `UNKNOWN` and excluded from comparative mess-level analysis.

A star-schema data warehouse was constructed to support analytical queries. The warehouse consists of a central fact table (`fact_expense`) containing transaction-level expenditure data, linked to dimension tables for date (`dim_date`), vendor (`dim_vendor`), and mess unit (`dim_mess_unit`). SQL queries were used to aggregate daily expenses, which were then analyzed using Python for trend identification and visualization.

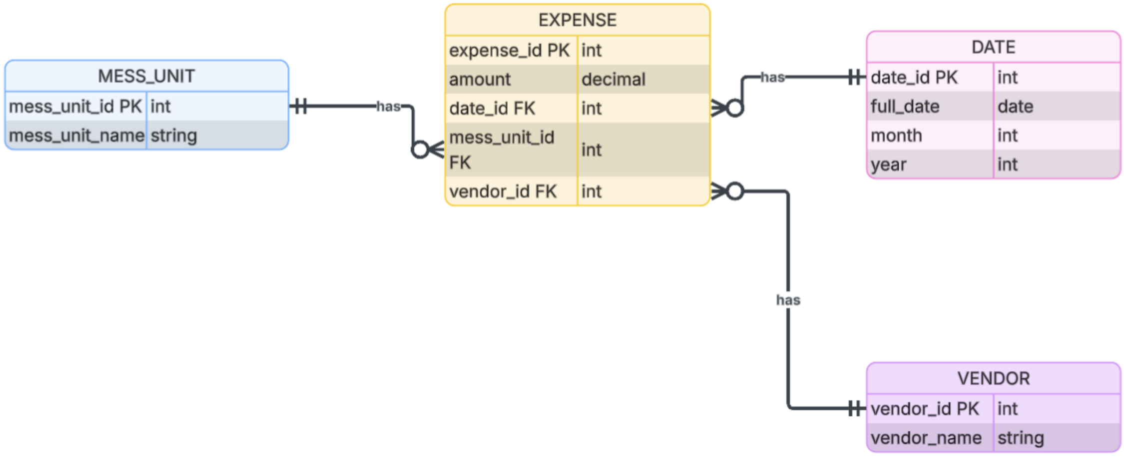


Figure 5.1: Entity-Relationship Diagram (ERD) for the database created.

Date	Total Expense (in Rs.)
2025-07-29	4,966.00
2025-08-01	250,787.80
2025-08-02	92,405.00
2025-08-03	31,216.10
2025-08-04	173,552.00
2025-08-05	438,681.70
2025-08-06	504,113.50
2025-08-07	252,638.90
2025-08-08	299,550.50
2025-08-09	96,592.50
2025-08-10	42,104.00
2025-08-11	493,431.80
2025-08-12	390,719.20
2025-08-13	35,992.30
2025-08-14	201,741.00
2025-08-15	219,135.55
2025-08-16	213,184.40
2025-08-17	39,432.50
2025-08-18	73,871.30
2025-08-19	287,289.60
2025-08-20	391,160.00
2025-08-21	116,897.30
2025-08-22	280,287.20
2025-08-23	51,847.50

The average daily expense for the month was calculated to be **Rs. 2,07,566.57**. Days with expenses exceeding this value were identified as *high-wastage risk days*, under the assumption that higher expenditure correlates with increased food preparation and procurement, thereby increasing the likelihood of food wastage.

Table 5.1: High-Wastage Risk Days (Expense Above Monthly Average)

Date	Total Expense (in Rs.)
2025-08-01	250,787.80
2025-08-05	438,681.70
2025-08-06	504,113.50
2025-08-07	252,638.90
2025-08-08	299,550.50
2025-08-11	493,431.80
2025-08-12	390,719.20
2025-08-15	219,135.55
2025-08-16	213,184.40
2025-08-19	287,289.60
2025-08-20	391,160.00
2025-08-22	280,287.20

For example, on 1st August 2025, multiple vendors such as vegetable suppliers, milk agencies, gas providers, and bakery vendors were paid separately. The reported daily expense of Rs. 250,787.80 for this date represents the sum of all individual vendor bills issued on 01-08-2025. During this aggregation process, no synthetic dates or artificial expense values were introduced; all daily totals are derived directly from the recorded transaction data.

Table 5.2: Estimated Food Wastage per Day Based on Expense Proxy

Date	Expense (in Rs.)	Estimated Wastage (kg)
2025-07-29	4,966.00	2.39
2025-08-01	250,787.80	120.82
2025-08-02	92,405.00	44.52
2025-08-03	31,216.10	15.04
2025-08-04	173,552.00	83.61
2025-08-05	438,681.70	211.35
2025-08-06	504,113.50	242.87
2025-08-07	252,638.90	121.71
2025-08-08	299,550.50	144.32
2025-08-09	96,592.50	46.54
2025-08-10	42,104.00	20.28
2025-08-11	493,431.80	237.72
2025-08-12	390,719.20	188.24
2025-08-13	35,992.30	17.34
2025-08-14	201,741.00	97.19
2025-08-15	219,135.55	105.57
2025-08-16	213,184.40	102.71
2025-08-17	39,432.50	19.00
2025-08-18	73,871.30	35.59
2025-08-19	287,289.60	138.41
2025-08-20	391,160.00	188.45
2025-08-21	116,897.30	56.32
2025-08-22	280,287.20	135.03
2025-08-23	51,847.50	24.98

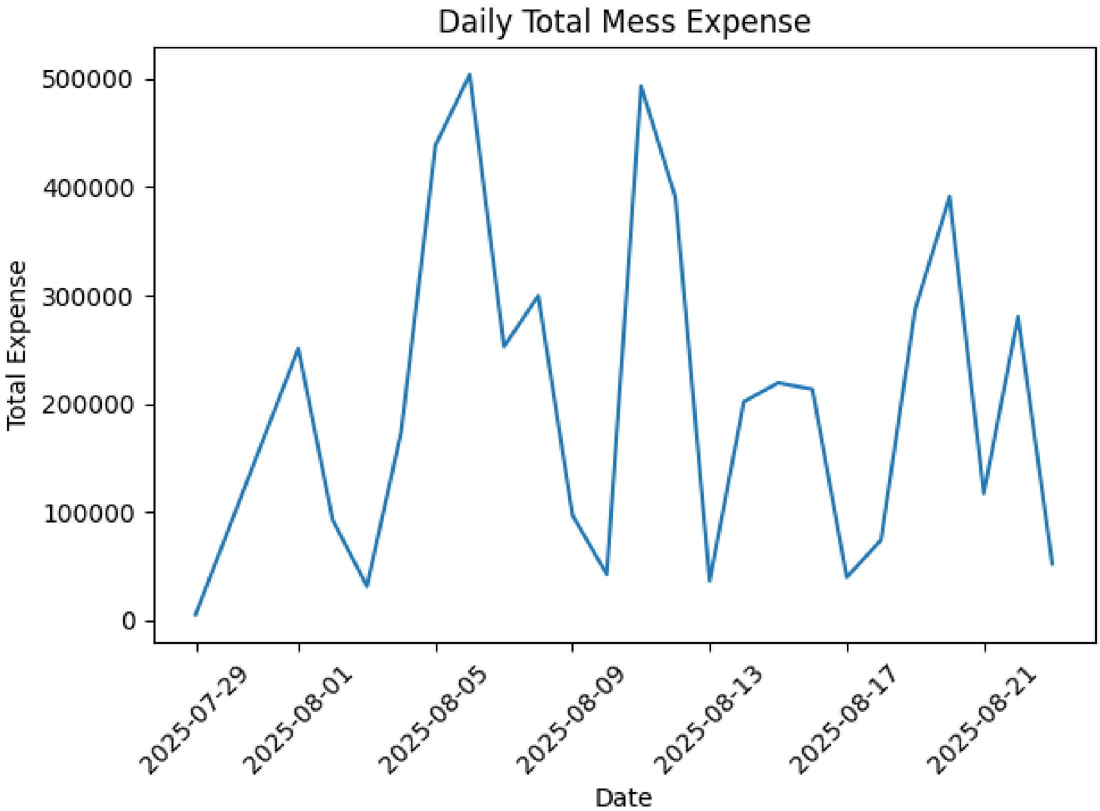


Figure 5.2: Daily Total Mess Expense Trend for August 2025.A pronounced spike in total expenditure is observed during the first two weeks of the semester, corresponding to the post–summer vacation influx of students.

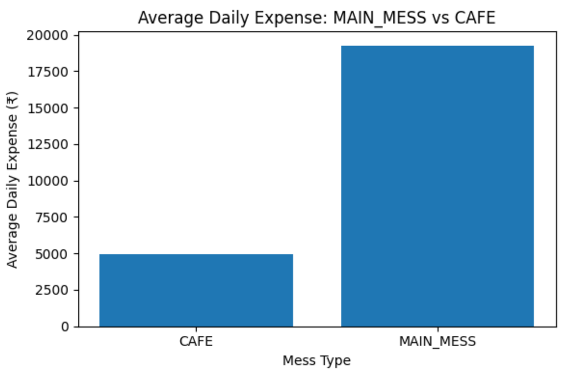


Figure 5.3: Average Daily Expense Comparison Between Main Mess and Café

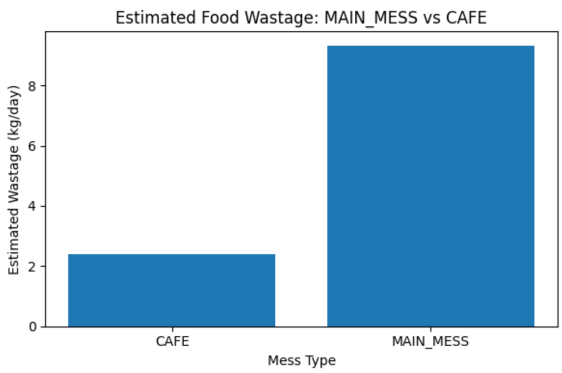


Figure 5.4: Estimated Food Wastage Comparison Between Main Mess and Café

For comparative performance analysis, CDH-1 and CDH-2 were combined into a single Main Mess category, while Café operations were analyzed separately. Transactions labeled as `UNKNOWN` were excluded. The comparison indicates that Café operations exhibit higher relative expenditure volatility and higher estimated wastage per unit expense compared to the Main Mess, reflecting irregular demand patterns and reduced predictability.

Chapter 6

September 2025 – The Month of Mid-Semester Examinations

September 2025 corresponds to a critical academic phase at IISER Thiruvananthapuram due to the commencement of mid-semester examinations around the third week of the month. This period is typically associated with changes in student activity patterns, extended study hours, increased café usage, and altered food consumption behavior. The following table presents the daily total mess expenditure and estimated wastage for each payment date in September 2025.

Table 6.1: Daily Expense with Estimated Food Wastage for September 2025

Date	Expense (Rs.)	Wastage (kg)
2025-09-01	202596.30	71.90
2025-09-02	1120119.60	397.50
2025-09-03	605340.60	214.82
2025-09-04	350151.90	124.26
2025-09-05	38046.00	13.50
2025-09-06	57508.50	20.41
2025-09-07	17056.50	6.05
2025-09-08	160001.40	56.78
2025-09-09	436371.90	154.86
2025-09-10	158896.50	56.39
2025-09-11	593793.40	210.72
2025-09-12	282928.90	100.40
2025-09-13	102785.70	36.48
2025-09-14	62354.50	22.13

Date	Expense (Rs.)	Wastage (kg)
2025-09-15	201242.30	71.42
2025-09-16	402985.05	143.01
2025-09-17	312381.60	110.86
2025-09-18	486320.80	172.58
2025-09-19	696649.90	247.22
2025-09-20	118013.50	41.88
2025-09-21	53465.00	18.97
2025-09-22	150539.40	53.42
2025-09-23	359252.30	127.49
2025-09-24	225701.90	80.10
2025-09-25	131142.60	46.54
2025-09-26	471503.20	167.32
2025-09-27	79349.40	28.16
2025-09-28	13656.50	4.85

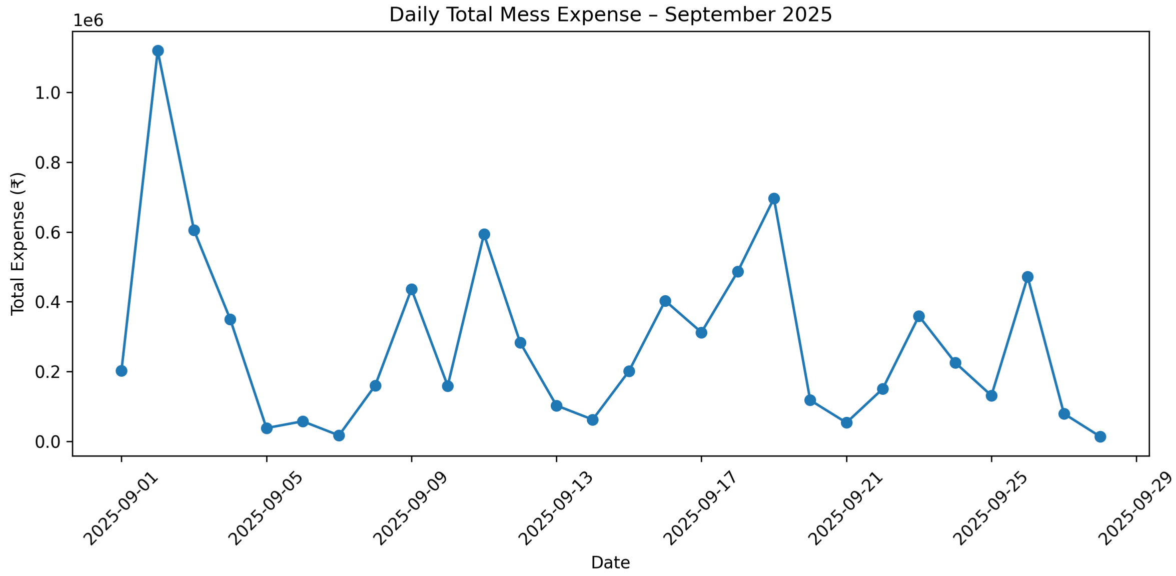


Figure 6.1: The spikes observed on 1st September and around mid-September can be explained by operational and behavioral factors. The increase on 1st September corresponds to the beginning of a new month, during which bulk procurement is typically undertaken to replenish inventory. The subsequent spike observed around mid-September aligns with the period preceding mid-semester examinations (commencing around 20th September). During this phase, students tend to shift their food consumption from the central mess to campus cafés due to extended study hours, leading to demand volatility and procurement–consumption mismatch within the mess.

The average daily expenditure for September 2025, computed directly from the daily totals listed above, was **Rs. 281,791.26**. Days on which the total mess expenditure exceeded this average were flagged as high-wastage-risk days, as elevated spending often correlates with over-procurement and increased food wastage. The dates identified through this criterion are presented alongside.Since direct food waste measurements were not available, waste was estimated using a proportional proxy model scaled against the monthly average daily expense. The table below lists the estimated food waste for each day of September 2025 exactly as computed by the Python analysis.

Table 6.2: High Wastage Days for September 2025

Date	Total Expense (in Rs.)
2025-09-02	1,120,119.60
2025-09-03	605,340.60
2025-09-04	350,151.90
2025-09-09	436,371.90
2025-09-11	593,793.40
2025-09-12	282,928.90
2025-09-16	402,985.05
2025-09-17	312,381.60
2025-09-18	486,320.80
2025-09-19	696,649.90
2025-09-23	359,252.30
2025-09-26	471,503.20

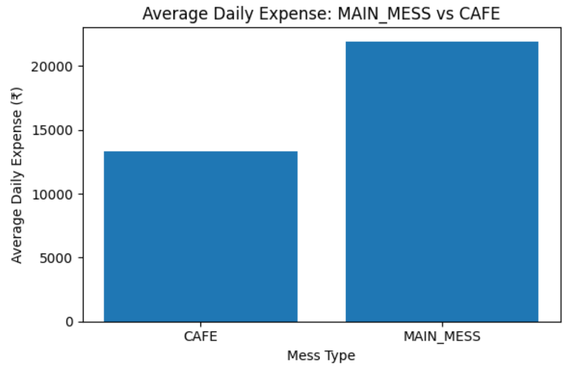


Figure 6.2: Average Daily Expense: Main Mess vs Café

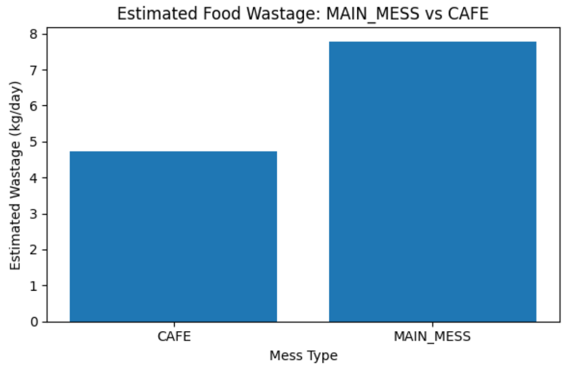


Figure 6.3: Estimated Food Wastage: Main Mess vs Café

Overall, the observed data trends align with our expectations. During the mid-semester examination period, there was an increase in food demand and irregular consumption patterns, consistent with anticipated student behavior. Figure 3.1 further illustrates the expected decrease in total mess expenditure following the conclusion of exams on 27th September, as many students returned home. A total of 390 records were analyzed to assess the mess expenses for September.

Chapter 7

October 2025 – The Month of Science and Sports Fests

October 2025 marks a distinct operational phase in the IISER Thiruvananthapuram mess ecosystem. Unlike September, which was dominated by academic pressures due to mid-semester examinations, October witnessed a sharp resurgence in campus activity driven by the return of students after semester breaks and the hosting of major academic and cultural events. From around 5^{th} October onwards, student footfall increased substantially, which is clearly reflected in the expenditure and estimated food wastage trends observed during the month.

The average daily mess expenditure for October 2025 was Rs. 265,018.59. Several days exhibited significantly higher-than-average expenses, particularly 2^{nd}, 4^{th}, 7^{th}, 10^{th}, 13^{th}, 14^{th}, 15^{th}, 16^{th}, 17^{th}, 20^{th}, and 24^{th} October. These spikes coincide with the timeline of major campus events. The dominance of the Main Mess in both expenditure and estimated food wastage remains consistent; however, the Café exhibits a noticeable rise in activity during this period. This aligns strongly with the timing of two major campus events — **ANVESHA**, one of the largest science festivals, and **ITSAV**, a multi-sport cultural event.

Table 7.1: October 2025 Mess Expenditure and High Wastage Days

Table 7.2: Daily Expense with Estimated Food Wastage for October 2025

Date	Expense (Rs.)	Wastage (kg)
2025-10-01	135524.60	51.14
2025-10-02	594604.60	224.36
2025-10-03	170413.70	64.30
2025-10-04	671767.00	253.48
2025-10-05	42931.00	16.20
2025-10-06	116067.40	43.80
2025-10-07	590569.05	222.84
2025-10-08	233508.90	88.11
2025-10-09	139453.50	52.62
2025-10-10	335743.50	126.69
2025-10-11	104014.30	39.25
2025-10-12	64322.70	24.27
2025-10-13	327927.70	123.74
2025-10-14	395754.80	149.33
2025-10-15	439174.40	165.71
2025-10-16	301777.90	113.87
2025-10-17	347779.60	131.23
2025-10-18	155347.00	58.62
2025-10-19	122469.00	46.21
2025-10-20	285274.90	107.64
2025-10-21	225528.00	85.10
2025-10-22	113787.50	42.94
2025-10-23	149616.50	56.46
2025-10-24	544229.60	205.36
2025-10-25	17877.60	6.75

Table 7.3: High Wastage Days for October 2025

Date	Expense (Rs.)
2025-10-02	594,604.60
2025-10-04	671,767.00
2025-10-07	590,569.05
2025-10-10	335,743.50
2025-10-13	327,927.70
2025-10-14	395,754.80
2025-10-15	439,174.40
2025-10-16	301,777.90
2025-10-17	347,779.60
2025-10-20	285,274.90
2025-10-24	544,229.60

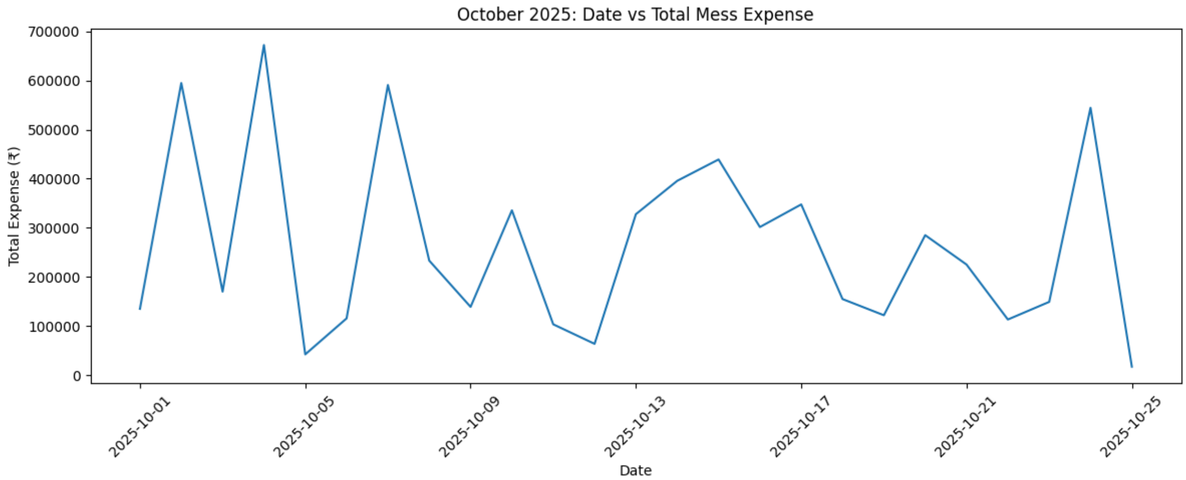


Figure 7.1: Date vs Total Mess Expenditure for October 2025

Table 7.4: Average Daily Expense and Estimated Wastage by Mess Group (October 2025)

Mess Group	Avg. Expense (in Rs.)	Avg. Estimated Wastage (kg)
CAFÉ	11066.743204	4.175837
MAIN_MESS	23244.026271	8.770715

Overall, October 2025 represents a month where expenditure patterns deviate from routine academic cycles and are instead driven by institutional events and student congregation dynamics. While the observed trends may appear intuitive, their quantitative confirmation through data-driven analysis is essential for validating operational assumptions and forms the foundation for more advanced predictive and optimization models in subsequent phases of this project.

Chapter 8

November 2025 – The Month of End Semester Examinations

November 2025 represents a crucial academic phase at IISER Thiruvananthapuram, marked by the commencement of end semester examinations from 19^{th} November onward. This transition is clearly reflected in the mess expenditure patterns observed during the month. As students returned to campus and academic intensity increased, both food consumption and associated operational expenses exhibited noticeable fluctuations. The following analysis presents a detailed examination of daily expenditure trends, estimated food wastage, and mess-wise consumption behavior for November 2025.

Using the predefined wastage conversion factor, estimated food wastage was calculated for each day based on total daily expenditure. Table 8.1 presents the full dataset, including every recorded date, total expense, and corresponding estimated food wastage in kilograms. An aggregation by mess group reveals a clear difference between Café and Main Mess consumption patterns. The Main Mess consistently accounts for a higher average daily expense and estimated food wastage, reflecting its role as the primary dining facility during academically intensive periods such as end semester examinations. Overall, November 2025 demonstrates a strong linkage between academic schedules and food consumption behavior. The alignment of expenditure peaks with examination periods reinforces the importance of incorporating academic calendars into operational planning and wastage mitigation strategies for institutional food systems.

Table 8.1: Daily Expense with Estimated Food Wastage (November 2025)

Date	Amount (Rs.)	Estimated Wastage (kg)
2025-11-01	212433.40	38.571766
2025-11-02	110274.20	20.022608
2025-11-03	1196157.20	217.187580
2025-11-04	1324937.60	240.570379
2025-11-05	255270.20	46.349691
2025-11-06	977196.20	177.430590
2025-11-07	700999.30	127.281215
2025-11-08	175488.20	31.863586
2025-11-09	234207.20	42.525259
2025-11-10	372503.80	67.635925
2025-11-11	801512.00	145.531417
2025-11-12	890479.00	161.685253
2025-11-13	638829.00	115.992885
2025-11-14	844426.20	153.323395
2025-11-15	288885.00	52.453168
2025-11-16	288598.60	52.401166
2025-11-17	468070.00	84.987985
2025-11-18	828691.00	150.466338
2025-11-19	202563.00	36.779587
2025-11-20	388846.00	70.603196
2025-11-21	849324.40	154.212767
2025-11-22	583272.20	105.905376
2025-11-23	34250.32	6.218868

Table 8.2: High Wastage Days for November 2025

Date	Total Expense (Rs.)
2025-11-03	1196157.20
2025-11-04	1324937.60
2025-11-06	977196.20
2025-11-07	700999.30
2025-11-11	801512.00
2025-11-12	890479.00
2025-11-13	638829.00
2025-11-14	844426.20
2025-11-18	828691.00
2025-11-21	849324.40
2025-11-22	583272.20

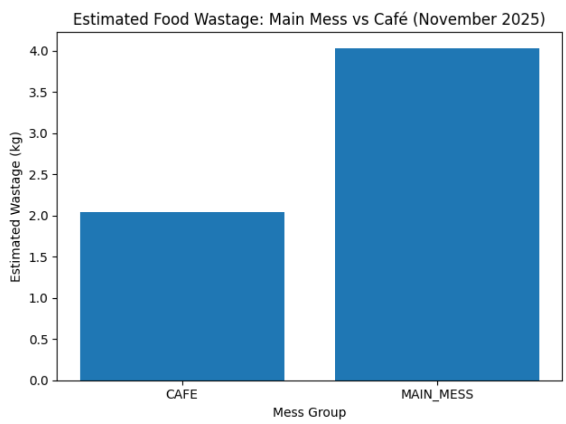


Figure 8.1: Estimated Food Wastage Comparison Between Main Mess and Cafe

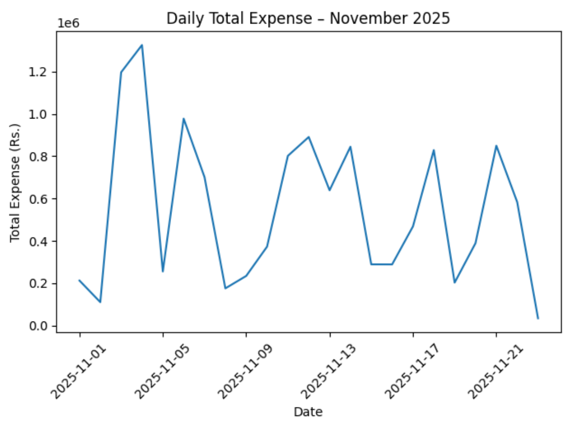


Figure 8.2: Daily Total Mess Expense for November 2025

Chapter 9

December 2025 — End of Semester and Campus Departure

December 2025 represents the final operational phase of the IISER Thiruvananthapuram student mess for the Varsha 2025 semester. Mess operations during this month were active only until 9th December, which was the last working day before the semester break. A pronounced decline in food expenditure is observed toward the end of the month, corresponding to the conclusion of end-semester examinations on 3rd December and the subsequent departure of students from campus.

The expense data for this month therefore captures a transition period, beginning with residual high demand immediately after examinations and ending with minimal activity as the campus population rapidly decreased.

Table 9.1: Daily Expense with Estimated Food Wastage — December 2025

Date	Expense (Rs.)	Estimated Wastage (kg)
2025-12-01	783448.10	346.56
2025-12-02	228160.10	100.93
2025-12-03	318818.00	141.03
2025-12-04	279594.70	123.68
2025-12-05	130693.20	57.81
2025-12-06	41159.20	18.21
2025-12-07	54994.20	24.33
2025-12-08	182834.60	80.88
2025-12-09	14898.00	6.59

The average daily expenditure for the operational days in December was Rs.**226,066.68**. The highest expenditure occurred on 1st December, reflecting continued campus occupancy immediately following examinations, while a sharp decline is visible after 5th December as students began vacating the campus.

Table 9.2: High Wastage Days — December 2025

Date	Total Expense (Rs.)
2025-12-01	783,448.10
2025-12-02	228,160.10
2025-12-03	318,818.00
2025-12-04	279,594.70

Table 9.3: Average Daily Expense and Estimated Wastage by Mess Group — December 2025

Mess Group	Avg. Daily Expense (Rs.)	Estimated Wastage (kg/day)
Café	13,990.68	6.19
Main Mess	18,401.02	8.14

These days fall immediately before and shortly after the end of examinations, a period characterized by uncertainty in student departure schedules and transitional consumption behavior.

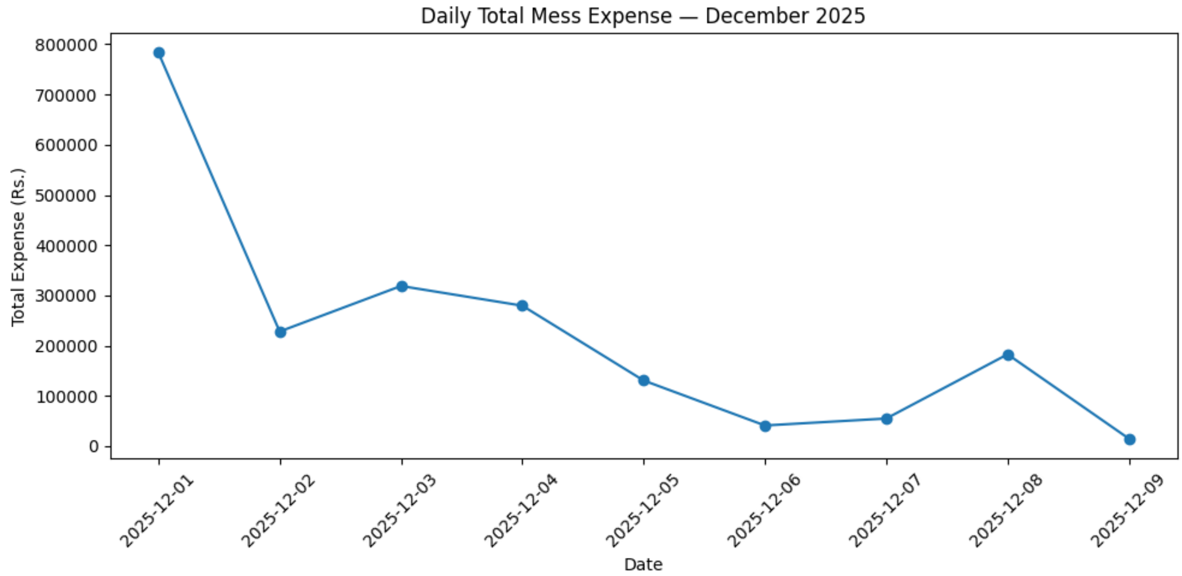


Figure 9.1: Daily Total Mess Expenditure — December 2025

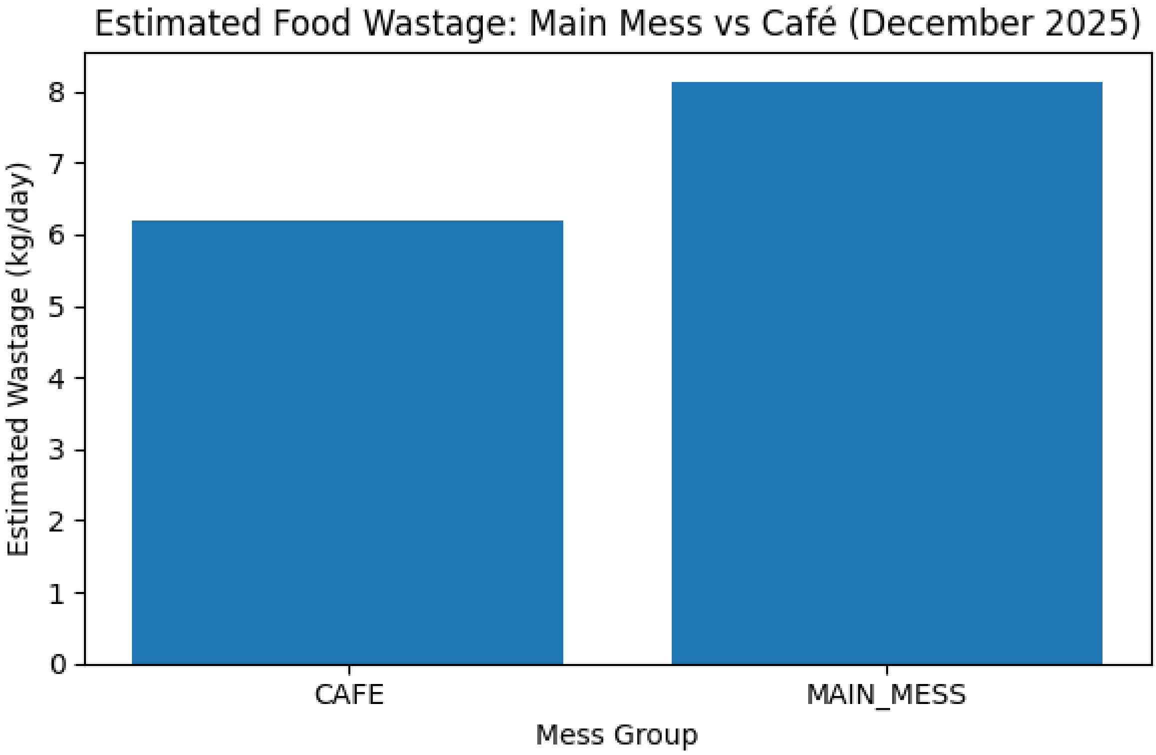


Figure 9.2: Estimated Food Wastage by Mess Group — December 2025

December exhibits a structurally different demand pattern compared to earlier months. Unlike spikes driven by academic events or festivals, expenditure dynamics here are dominated by campus attrition. The Main Mess shows higher average wastage than the Café, reflecting the difficulty of rapidly scaling down bulk meal preparation in response to sudden population decline. Café operations, being more demand-driven and flexible, adapt more efficiently during such transition periods.

The December 2025 analysis highlights the importance of adaptive procurement strategies during semester transitions. The observed decline in expenditure and wastage after early December confirms that demand forecasting during exit phases is as critical as during peak academic activity. Incorporating academic calendar milestones into procurement planning can substantially reduce food wastage during semester-end periods.

Chapter 10

Benford's Law Analysis of Mess Expenses

While monthly trend analysis provides valuable insights into operational behavior, business intelligence systems are most effective when they enable deeper diagnostic and forensic analytics. To extend this study beyond descriptive reporting, Benford's Law was applied to vendor-level mess expense data for Varsha 2025 as a preliminary fraud and anomaly detection technique.

Benford's Law, also known as the First-Digit Law, describes the empirical observation that in many naturally occurring numerical datasets, the leading digits are not uniformly distributed. Instead, smaller digits appear disproportionately often. The probability that a number has first digit $d \in \{1, 2, \ldots, 9\}$ is given by the logarithmic relationship:

$$P(d) = \log_{10}\left(1 + \frac{1}{d}\right)$$

This results in the digit 1 appearing as the leading digit approximately 30.1% of the time, while higher digits occur with progressively lower frequencies. This phenomenon has been extensively validated across financial, demographic, and economic datasets and is widely used in forensic accounting and audit analytics.

The applicability of Benford's Law relies on datasets that span multiple orders of magnitude and arise from natural economic processes rather than artificially assigned values. Vendor-level expense totals satisfy these criteria, as they are generated through repeated procurement transactions involving diverse suppliers and varying order sizes. Consequently, Benford's Law serves as a suitable screening tool for the mess expense data warehouse constructed in this project.

For this analysis, vendor-wise total expenses were aggregated across the Varsha 2025 period using the data warehouse fact and dimension tables. The leading digit of each vendor's cumulative expense was extracted, and the observed frequency distribution was compared against the theoretical Benford distribution. Figure 10.2 presents this comparison.

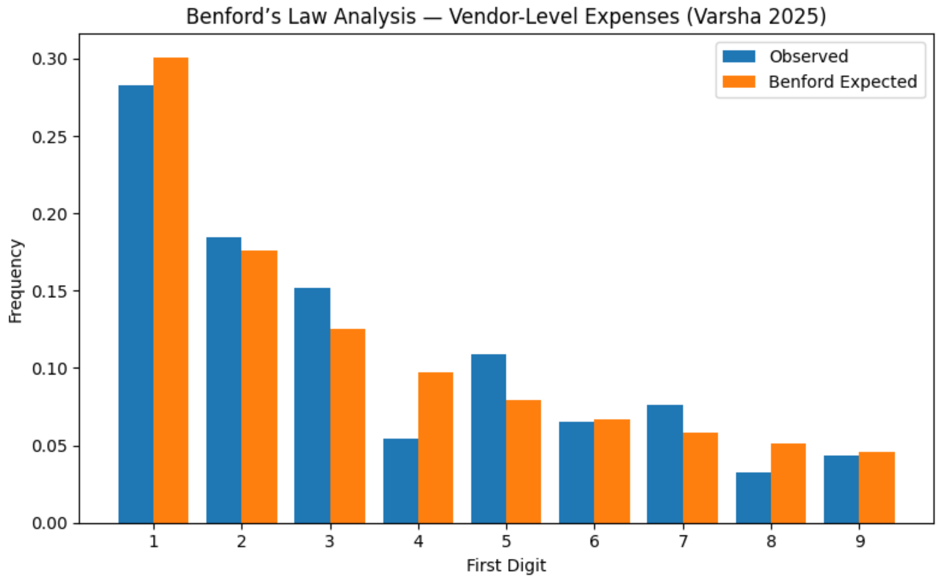


Figure 10.1: Benford's Law Analysis of Vendor-Level Mess Expenses for Varsha 2025

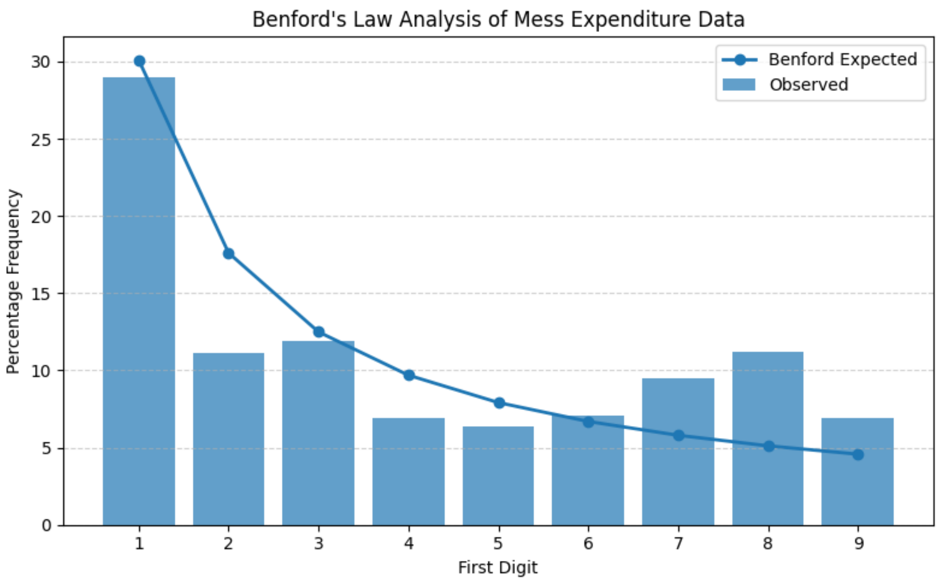


Figure 10.2: Observed vs Benford-Expected Leading Digit Distribution for Mess Expenditure Data

Methodology

The analytical workflow consisted of the following steps:

1. Extraction of vendor-level expenditure values from the data warehouse.
2. Removal of zero or null values to avoid undefined leading digits.
3. Extraction of the first non-zero digit from each transaction amount.
4. Computation of observed frequency counts for digits 1 through 9.
5. Calculation of expected frequencies using Benford's Law.
6. Statistical comparison using a chi-square goodness-of-fit test.

Statistical Validation: Chi-Square Goodness-of-Fit Test

To quantitatively assess the deviation, a chi-square goodness-of-fit test was performed at a 5% significance level.

The test statistic is defined as:

$$\chi^2 = \sum_{d=1}^{9} \frac{(O_d - E_d)^2}{E_d} \tag{10.1}$$

where:

- O_d is the observed count for digit d
- E_d is the expected count based on Benford's Law = Total transaction entries $\cdot P(d) = 1829 \cdot P(d)$

Test Results

- Chi-square statistic: 261.70
- Degrees of freedom: 8
- Critical value at $\alpha = 0.05$: 15.51
- P-value: < 0.001

Since the computed chi-square statistic greatly exceeds the critical value, the null hypothesis that the data follows Benford's Law is rejected at the 5% significance level.

Observed and Expected Frequencies

Table 10.1 presents the detailed comparison between observed and expected digit frequencies.

Table 10.1: Observed vs Expected Leading Digit Frequencies

Digit	Observed Count	Expected Count	Observed (%)	Expected (%)
1	530	550.58	28.98	30.10
2	203	322.07	11.10	17.61
3	218	228.51	11.92	12.49
4	127	177.25	6.94	9.69
5	117	144.82	6.40	7.92
6	130	122.45	7.11	6.69
7	173	106.07	9.46	5.80
8	205	93.56	11.21	5.12
9	126	83.69	6.89	4.58

Limitations

It is important to acknowledge that Benford's Law is sensitive to data generation mechanisms. Institutional procurement systems may violate some assumptions of randomness due to negotiated pricing, fixed-rate contracts, and batch payments. Therefore, Benford's Law should be interpreted as an exploratory diagnostic tool rather than definitive proof of irregularity.

Conclusion

The application of Benford's Law to vendor-level mess expenditure data provides an additional analytical dimension beyond descriptive statistics and trend analysis. By combining visual inspection with formal statistical testing, this study demonstrates how forensic data mining techniques can be integrated into data warehousing and business intelligence workflows to enhance financial transparency and audit readiness.

Chapter 11

Network Analysis of the Mess–Vendor Financial System

This chapter presents a network-based analysis of the IISER Thiruvananthapuram student mess ecosystem using transaction data from Varsha 2025. While earlier chapters focused on temporal expenditure trends and aggregate financial behavior, the objective here is to uncover the *structural organization* of financial relationships between mess units and vendors. By modeling the system as a weighted graph, this chapter identifies dependency patterns, financial hubs, and naturally emerging communities within the mess procurement network.

Network Construction from the Data Warehouse

The mess–vendor financial network was constructed directly from the data warehouse by integrating information from the fact and dimension tables. Each node in the network represents either a mess entity (CDH-1, CDH-2, Café, and UNKNOWN) or an individual vendor. An undirected edge exists between a vendor and a mess unit if at least one financial transaction occurred between them during Varsha 2025. The weight assigned to each edge corresponds to the **total monetary value of transactions** between that vendor and mess unit over the academic year.

This modeling approach transforms transactional data into a relational structure, allowing financial dependencies to be analyzed as network properties rather than isolated records. The resulting network contains **96 nodes** and **114 edges**, capturing the complete vendor–mess interaction landscape without aggregation or pruning.

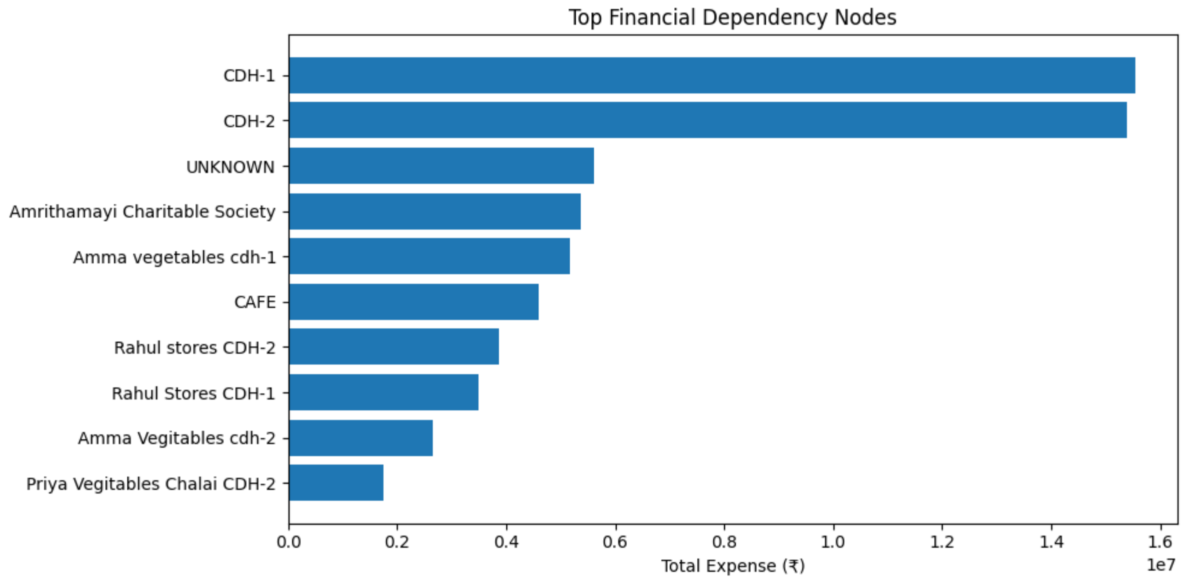
Global Network Characteristics

The global structure of the network exhibits a hub-and-spoke topology, where mess units act as central hubs connected to multiple vendors. These nodes serve as the primary financial anchors of the system, while several vendors remain sparsely connected, reflecting low-frequency or specialized procurement relationships .A notable feature of the network is the presence of the *UNKNOWN* node, which aggregates expenses such as cash vouchers and administrative transactions. Its relatively high connectivity highlights the importance of improving expense classification and auditing mechanisms within the financial workflow.

Centrality and Financial Dependency Analysis

To quantify structural importance, degree centrality and weighted degree were computed for each node. Degree centrality measures the number of direct connections a node has, while weighted degree captures the total financial volume flowing through that node.

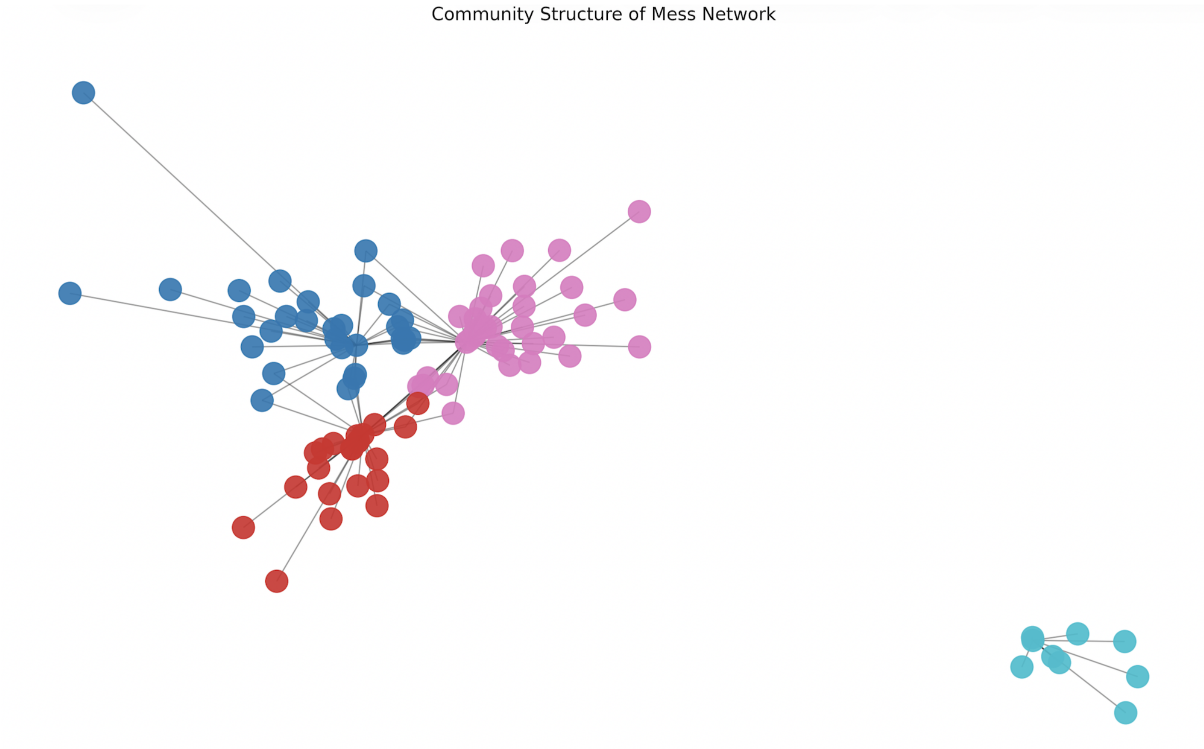
The mess units CDH-1 and CDH-2 emerge as the most dominant nodes, with weighted degrees of Rs.15.54 million and Rs 15.38 million respectively. Café also exhibits high degree centrality, reflecting consistent interactions with multiple vendors. Several vendors, including *Amrithamayi Charitable Society*, *Amma Vegetables*, and *Rahul Stores*, appear as significant financial dependency points due to sustained high-value transactions with specific mess units.



These high-weight vendors represent potential *risk concentration points*, where operational or pricing disruptions could disproportionately impact mess operations.

Community Detection and Modularity

To uncover latent organizational structure, community detection was performed using a modularity-maximization algorithm. Importantly, the number of communities was **not predefined**. Instead, the algorithm autonomously partitioned the network into **four communities**, achieving a modularity score of **0.650**. Such a high modularity value indicates strong separation between communities, meaning nodes are far more densely connected within communities than across them.



Conclusion

This chapter illustrates how network analysis complements traditional business intelligence techniques by revealing relational and structural properties that remain invisible in aggregate or time-series analyses. Using the data warehouse and applying graph-theoretic methods, the mess–vendor system was shown to have clear financial hubs, distinct communities, and identifiable risk concentration points. These findings provide a solid foundation for advanced analyzes such as procurement optimization, anomaly detection, and policy-driven restructuring of vendor dependencies.

Chapter 12

Codes and Datasets

All datasets and source code used in this study are uploaded on Github with this submission.