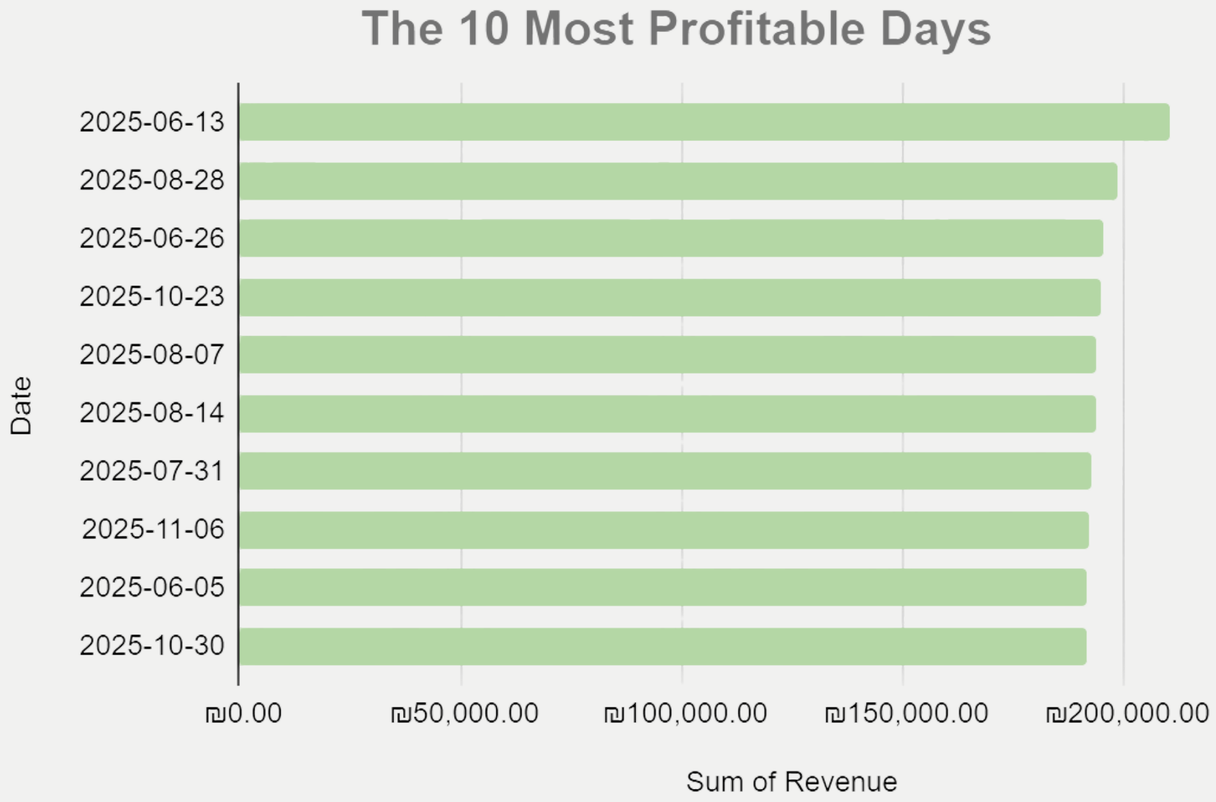
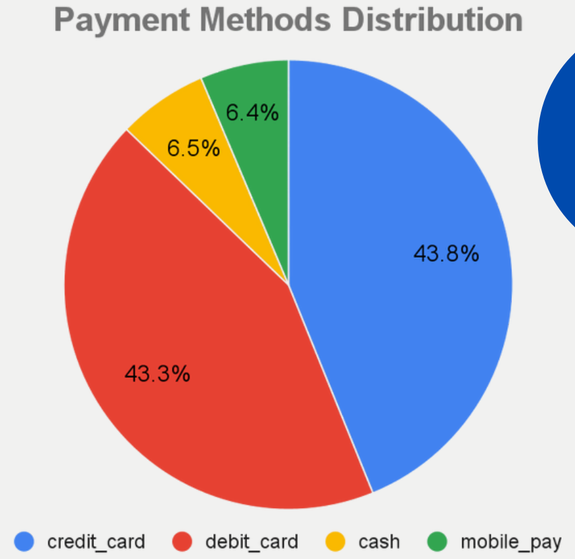
# Retail Data Analysis Project

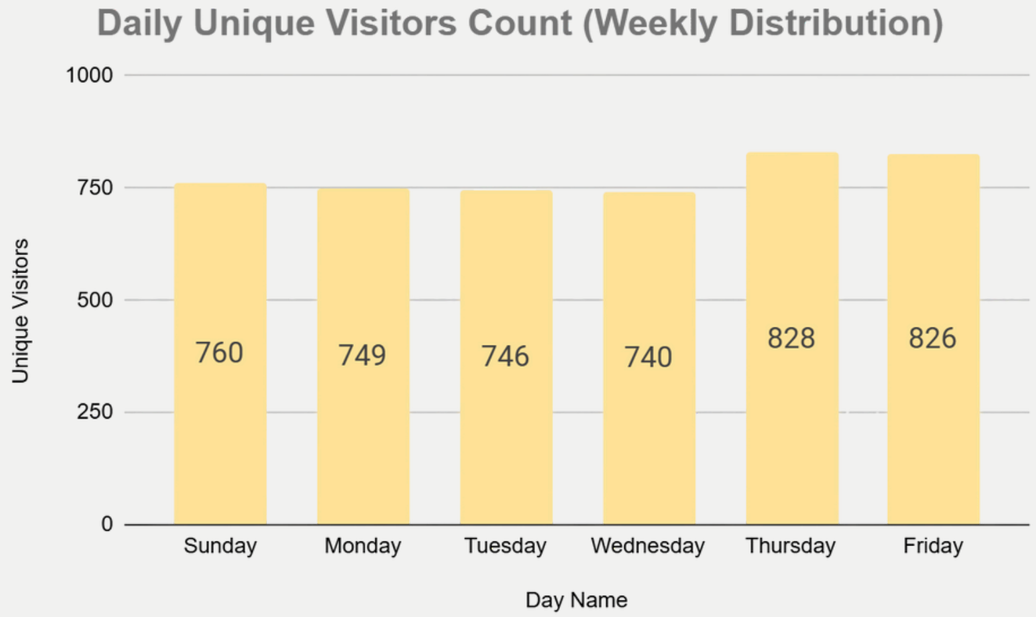
Final Project - Data Analyst Program
Google and Reichman AI Tech School

**Dor Thaler**

# EDA

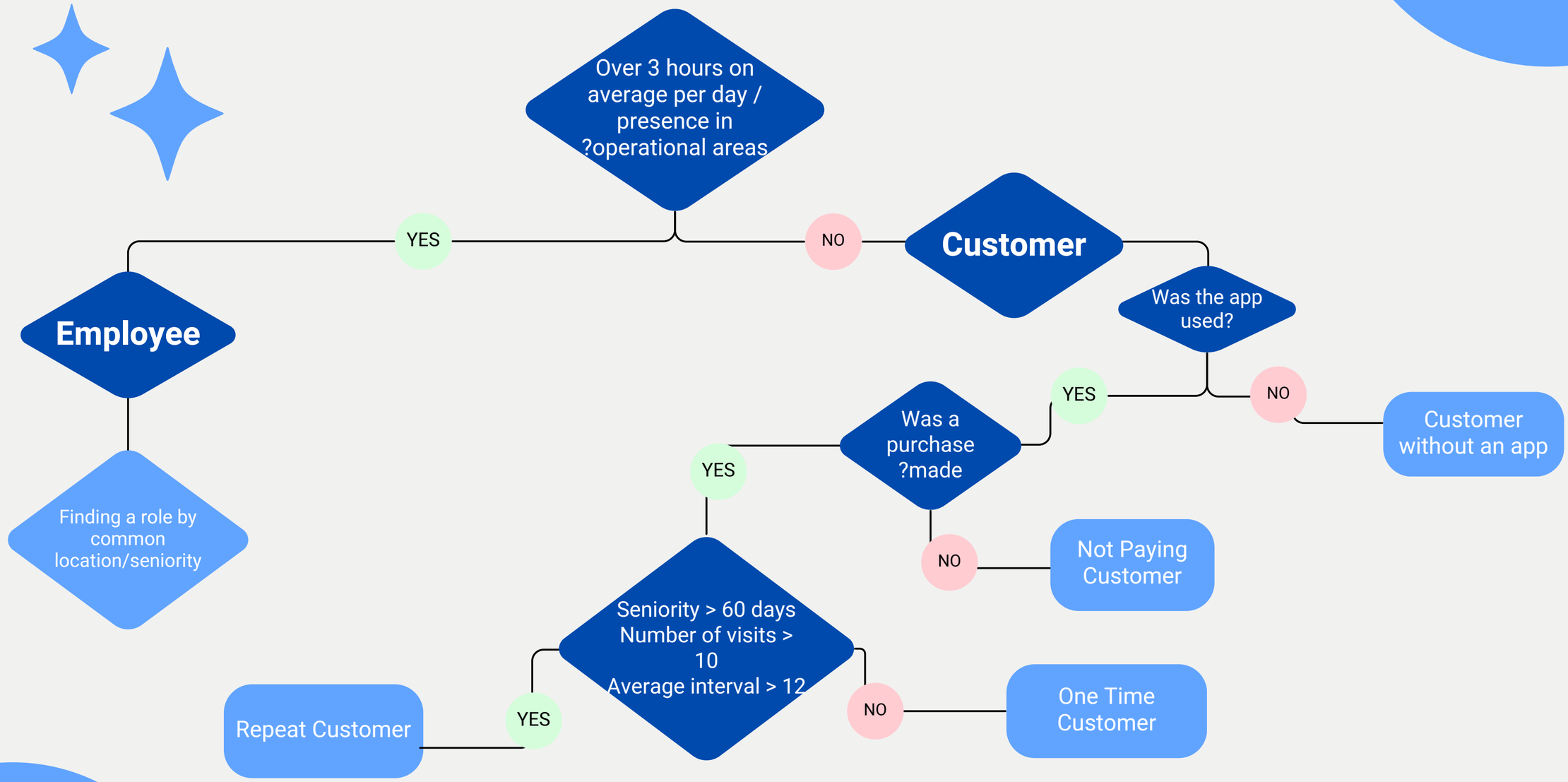



**The most profitable day is 13/6/25**
**The day the Israel-Iran war began**

# Classification of employees and customers



# Identifying outliers with location accuracy

## Main Insights:

Total signals: 1.3M, abnormal signals: 415K (over 30m).

This is an abnormal rate of about 32% (about a third of the data).

- Security guards have the highest abnormal rate, with about 89% of the signals defined as having low accuracy

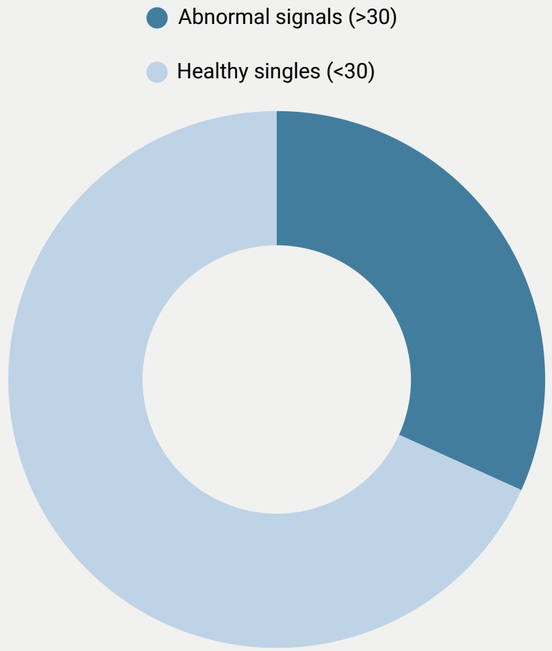
## Data processing method:

Differential Method:

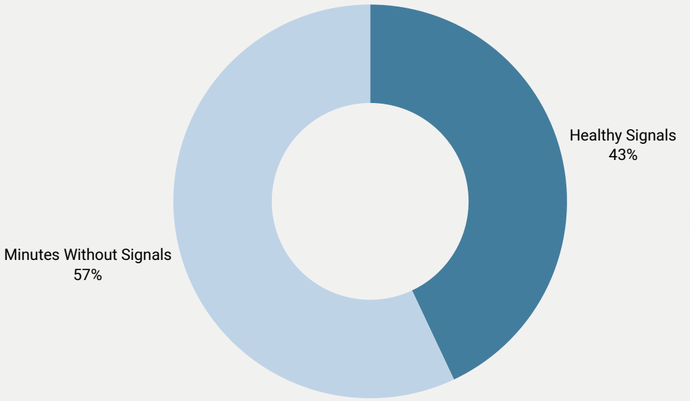
Static users (e.g. cashier, butcher): Data completion

For dynamic users (customers and general employees):

- In location-based analysis: Filtering and omitting exceptions
- In general attendance analysis: Retention of records, even in case of incorrect location.



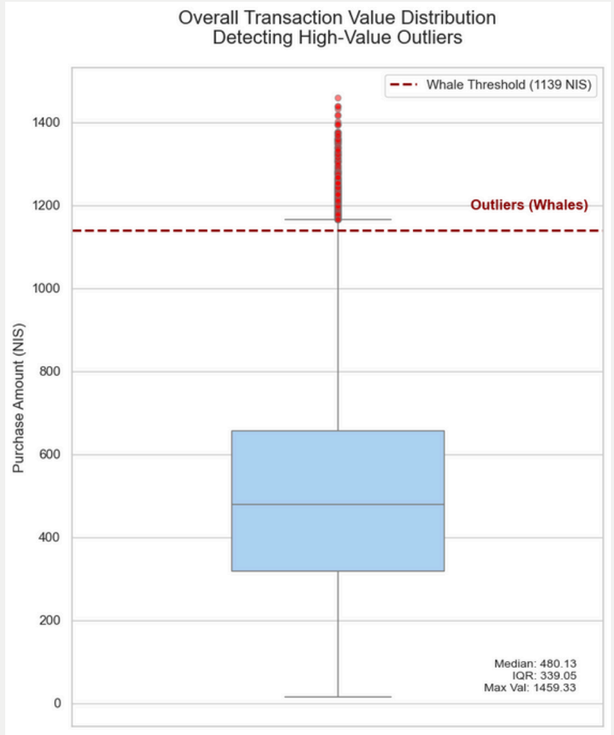
# Application Performance Testing



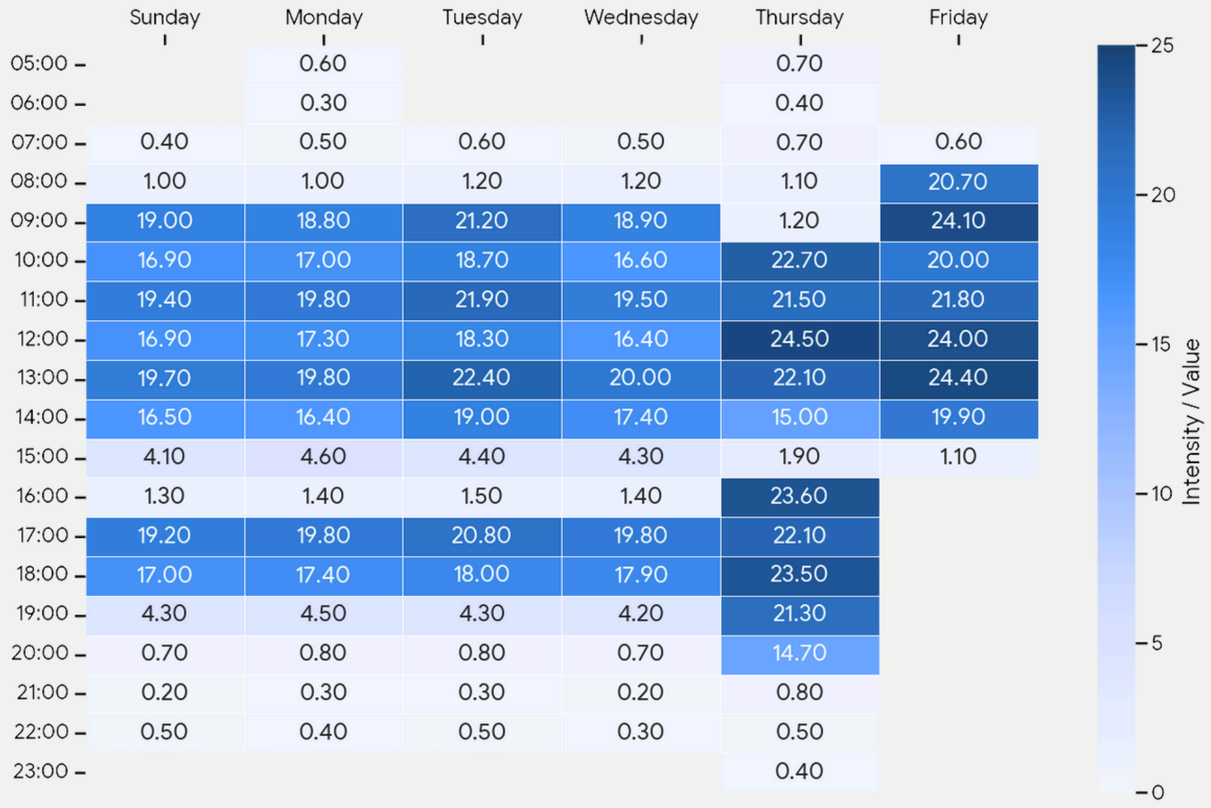

- Algorithm for daily supermarket sessions for each customer with a dedicated app ('Sessionization').
- Algorithm that isolates separate visits and calculates the net dwell time with maximum accuracy.
- Weighted score is 43% - The app meets the threshold requirements

# Unusual Transactions

- Using an IQR statistical test, 334 transactions were identified that deviated drastically from the norm (over 1,139 NIS).
- The data segmentation showed that 100% of these transactions belonged to repeat customers.
- Main Insight - Proposal for further segmentation by average basket size (Small, Medium, Large)



# Business Recommendations

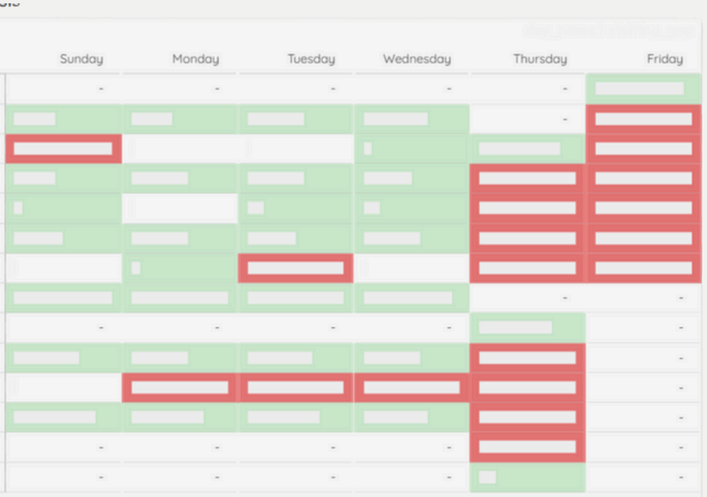



- Consistent and sharp drop in activity every day around 3:00 PM
- The maximum load is concentrated on Thursday evening (5:00 PM-8:00 PM) and on Friday
- The almost complete dominance of the Medium segment: about 99% of sales during peak and off-peak hours are made by customers with a shopping basket of 300–800 NIS. In a homogeneous audience of families who make routine weekly purchases.
- Recommendations: “Traffic light”, and Demand Shifting of the Medium segment through discounts, points in the app and promotions during off-peak hours.

# Optimization of supermarket checkouts

## recommendation for further development

**Gap Analysis - Calculating gaps in the number of available checkouts versus the average load at checkouts. Based on analysis of customer geolocation data and sales history.**




## Conclusions:

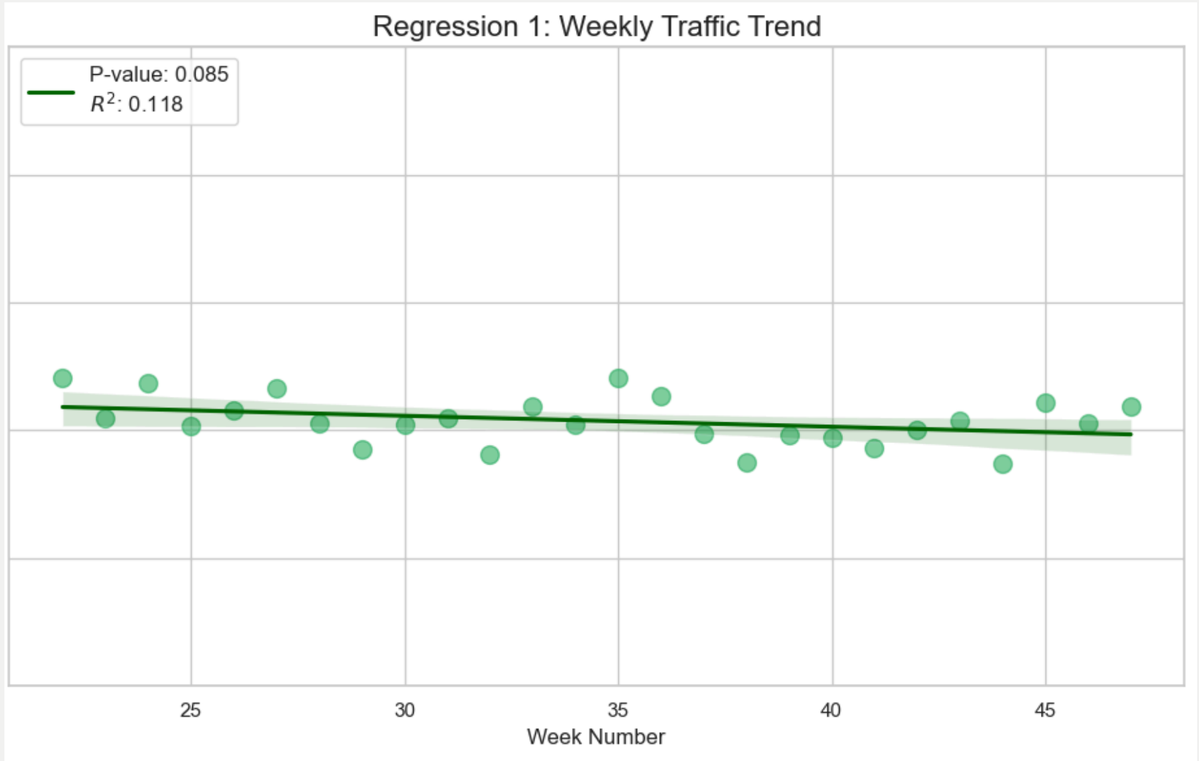
- The average cashier's handling time per customer is 1.5 - 1.25 minutes.
- There are gaps in peak versus off-peak times, between the number of customers and the number of cash registers available.

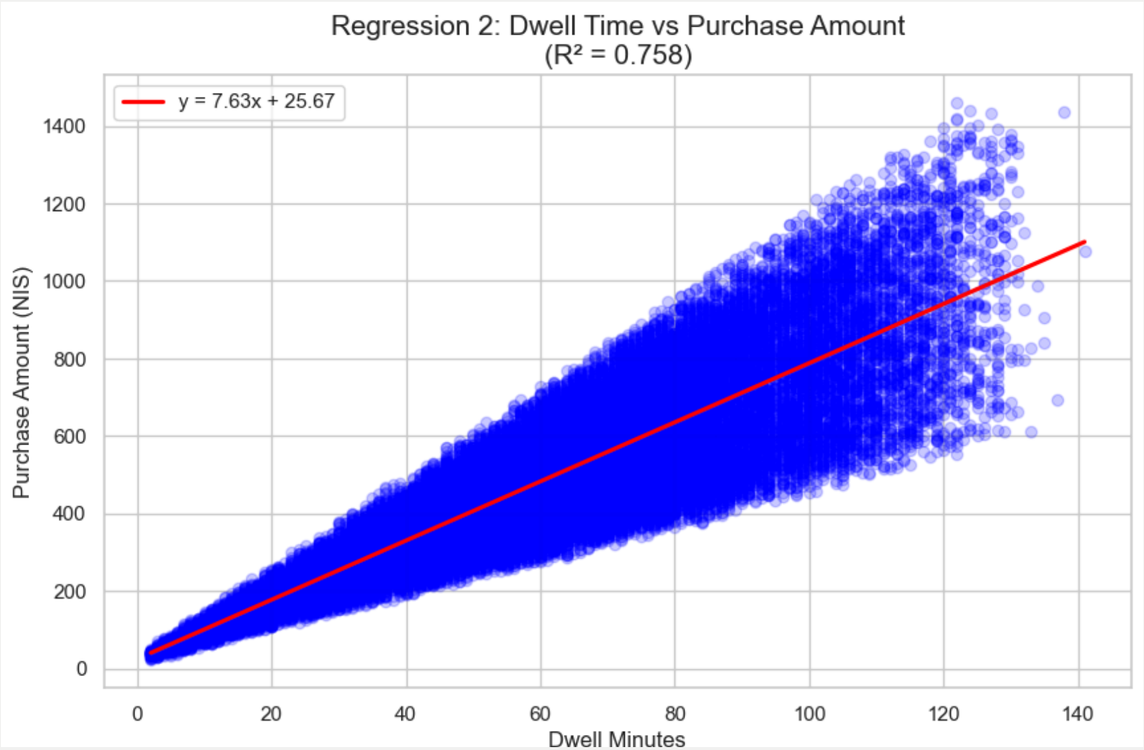
## Recommendations:

- Introducing self-service checkouts and the Scan & Go app - make it possible to cope with the high load.
- During off-peak hours - preventing "hidden unemployment" of cashiers (reducing shifts or efficient use of manpower).

# Regression Analysis 1

Visitor Trend over Time




# Regression Analysis 2

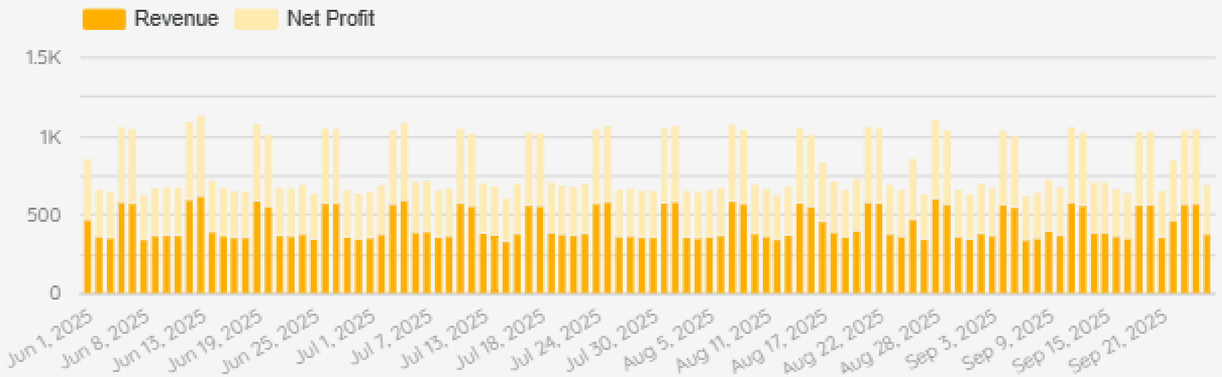
Purchase amount depending on the length of stay

# Financial & Sales Performance

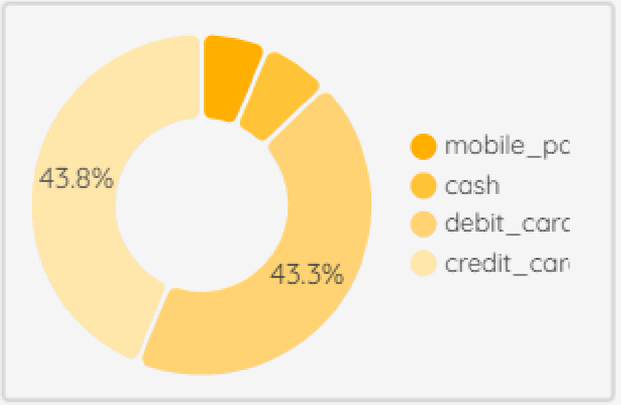
Select date range

| Total Sales | Total Revenue | Net Profit | Average Purchase | New Customers | Retention Rate |
|---|---|---|---|---|---|
| 34.6K | 16.55M₪ | 14.03M₪ | 478₪ | 1.04K | 99.83% |

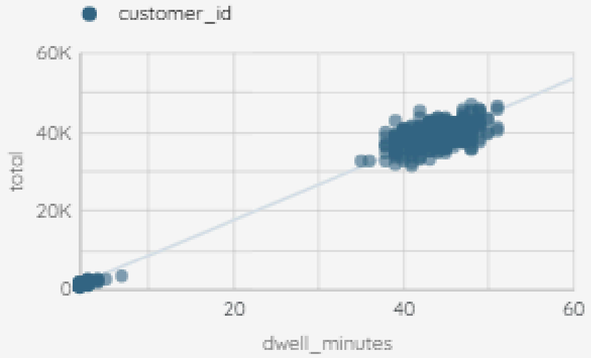
## Revenue & Net Profit



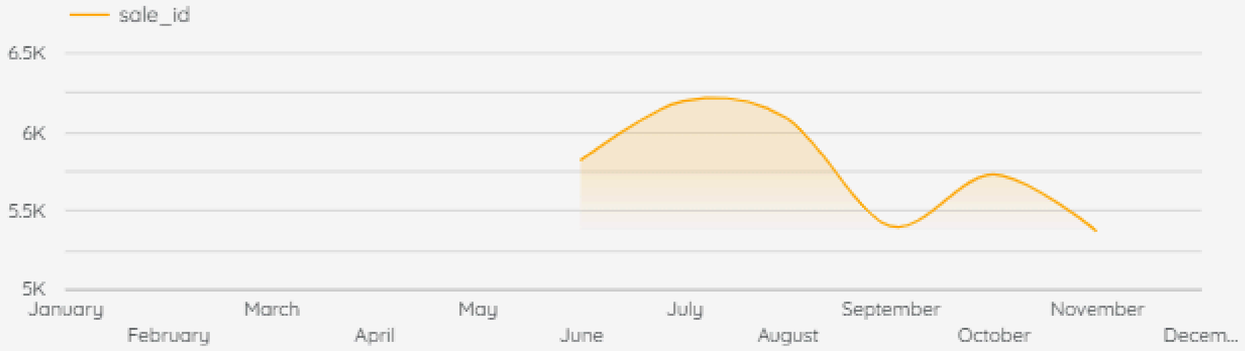

## Payment Method Distribution




## Purchase Amount vs. Dwell Time




## Sales Per Month



# Store Traffic & Operations

Oct 1, 2025 - Dec 27, 2025

| App Users | New Customers | Past Month Total Traffic | App Conversion Rate | Staffing Fit Ratio | Peak Wait Time |
|---|---|---|---|---|---|
| 844 | 54 | 452 | 0.67% | 91% | 92 Min |
| -8.7% | 1,700.0% | | -3.9% | -0.9% | 14.3% |

## Traffic Heatmap

| Hour | Sunday | Monday | Tuesday | Wednesday | Thursday | Friday |
|---|---|---|---|---|---|---|
| 8AM | 1 | 1 | 1 | 1 | 1 | 21 |
| 9AM | 19 | 19 | 21 | 19 | 1 | 24 |
| 10AM | 17 | 17 | 19 | 17 | 23 | 20 |
| 11AM | 19 | 20 | 22 | 20 | 22 | 22 |
| 12PM | 17 | 17 | 18 | 16 | 25 | 24 |
| 1PM | 20 | 20 | 22 | 20 | 22 | 24 |
| 2PM | 17 | 16 | 19 | 17 | 15 | 20 |
| 3PM | 4 | 5 | 4 | 4 | 2 | 1 |
| 4PM | 1 | 1 | 2 | 1 | 24 | - |
| 5PM | 19 | 20 | 21 | 20 | 22 | - |
| 6PM | 17 | 17 | 18 | 18 | 24 | - |
| 7PM | 4 | 5 | 4 | 4 | 21 | - |
| 8PM | 1 | 1 | 1 | 1 | 15 | - |
| 9PM | 0 | 0 | 0 | 0 | 1 | - |
| 10PM | 1 | 0 | 1 | 0 | 1 | - |

## Gap Analysis

day_name / staffing_gap

| hour | Sunday | Monday | Tuesday | Wednesday | Thursday | Friday |
|---|---|---|---|---|---|---|
| 8AM | - | - | - | - | - | |
| 9AM | | | | | - | |
| 10AM | | | | | | |
| 11AM | | | | | | |
| 12PM | | | | | | |

## Outstanding employees

| Work | role | Appearance |
|---|---|---|
| sen_001 | senior_general_worker | |
| but_001 | butcher | |
| sec_003 | security_guy | |
| but_004 | butcher | |
| gen_008 | general_worker | |

# Thank You!