# FoodCorp – Data Analysis
## 20535493

## Data Cleaning Steps:

Prior to running KPI queries, a thorough analysis of the given tables was conducted:

- Null values and instances of zero quantities were identified and subsequently removed.
- Receipt lines lacking clear product codes or having undefined departments were deleted.
- Unusually high-priced commodities in specific receipts were identified. Two receipt lines with such anomalies had their product values adjusted to the average value of that product.
- A new column called 'Store Loc' was added to STORES table to uniquely identify store names instead of codes. It is a combination of City and first part of postal code. Ex: Nottingham NG1.
- SQL codes were implemented to execute the data cleaning operations outlined above. These codes are made available at the end of Section 1.

## Section 1: The KPIs

### KPI 1: Total Sales per customer base

**KPI Description (in words):** Total sales in each store per customer base for a given period

**KPI formula:** sum(value)/count(unique customers), per quarter, per store

**Steps to realize KPI:**
1) Write the query below and load the extracted output as data source in Tableau

```
SELECT
   STORE_LOC,
   CONCAT('Q',EXTRACT(QUARTER FROM PURCHASED_AT), ' ',EXTRACT(YEAR FROM
PURCHASED_AT)) AS SALE_QUARTER,
   SUM(VALUE)/COUNT(DISTINCT R.CUSTOMER_ID) AS TOTAL_SALES_PER_CUST
FROM
   RECEIPT_LINES RL,
   RECEIPTS R,
   STORES S
WHERE R.RECEIPT_ID = RL.RECEIPT_ID
AND R.STORE_CODE = S.STORE_CODE
GROUP BY
   STORE_LOC, CONCAT('Q',EXTRACT(QUARTER FROM PURCHASED_AT), ' ',EXTRACT(YEAR FROM
PURCHASED_AT))
ORDER BY
   STORE_LOC, CONCAT('Q',EXTRACT(QUARTER FROM PURCHASED_AT), ' ',EXTRACT(YEAR FROM
PURCHASED_AT));
```

2) Visualized via Tableau as graph titled "KPI 1". See the Tableau file.
3) Data underpins Figure 1 in the comparative analysis.

**Additional Notes:** Quarters are represented by concatenating 'Q' with quarter number followed by year. Total sales per customer is calculated as the sum of receipt totals for a store divided by the unique number of customers shopped in that store.

**KPI 2: Customer Lifetime Value**

**KPI Description (in words):** Customer Lifetime Value

**KPI formula:** Average(value), per customer, per store

**Steps to realize KPI:**
1) Write the query below and load the extracted output as data source in Tableau

```
SELECT STORE_LOC,
AVG(CUST_SPEND) AS AVG_CUST_SPEND
FROM
(SELECT S.STORE_LOC, R.CUSTOMER_ID, SUM(RL.VALUE) AS CUST_SPEND
FROM RECEIPTS R,
RECEIPT_LINES RL,
STORES S
WHERE R.RECEIPT_ID = RL.RECEIPT_ID
AND S.STORE_CODE = R.STORE_CODE
GROUP BY S.STORE_LOC, R.CUSTOMER_ID
)
GROUP BY STORE_LOC
ORDER BY 1
```

2) Visualized via Tableau as graph titled "KPI 2". See the Tableau file.
3) Data underpins Figure 2 in the comparative analysis.

**Additional Notes:** Customer lifetime value is calculated as the total amount a customer spends in the store over the entire time he/she was shopping. Mean value is taken as a reference for entire customer base per store.

**KPI 3: New Customers in each store per quarter**

**KPI Description (in words):** New customers acquired per quarter in each store

**KPI formula:** Count(customers), per quarter(min(purchased at)), per store

**Steps to realize KPI:**
1) Write the query below and load the extracted output as data source in Tableau

```
SELECT STORE_LOC, YEAR, QTR, COUNT(*) NEW_CUSTOMERS
FROM (
SELECT CUSTOMER_ID,
S.STORE_LOC,
YEAR(MIN(PURCHASED_AT)) AS YEAR,
CONCAT('Q',QUARTER(MIN(PURCHASED_AT))) QTR
FROM RECEIPTS R,
STORES S
WHERE R.STORE_CODE = S.STORE_CODE
GROUP BY 1, 2
)
GROUP BY 1,2,3
ORDER BY 1,2,3
```

2) Visualized via Tableau as graph titled "KPI 3". See the Tableau file.
3) Data underpins Figure 3 in the comparative analysis.

**Additional Notes:** Customer is considered newly acquired in a specific quarter if their first purchase occurs in that quarter.

**KPI 4: Habitual Customers**

**KPI Description (in words):** Regular/Habitual customers per quarter in each store

**KPI formula:** Count(unique weeks), per customer, per quarter, per store

**Steps to realize KPI:**
1) Write the query below and load the extracted output as data source in Tableau

```
WITH CustomerWeeklyCounts AS (
    SELECT
        STORE_CODE,
        EXTRACT(YEAR FROM PURCHASED_AT) AS YEAR,
        EXTRACT(QUARTER FROM PURCHASED_AT) AS QUARTER,
        CUSTOMER_ID,
        COUNT(DISTINCT DATE_TRUNC('WEEK', PURCHASED_AT)) AS WEEKLY_COUNT
    FROM
        RECEIPTS
    GROUP BY
        STORE_CODE, YEAR, QUARTER, CUSTOMER_ID
)

SELECT
    S.STORE_LOC,
    C.YEAR,
    C.QUARTER,
    COUNT(DISTINCT C.CUSTOMER_ID) AS HABITUAL_CUSTOMERS
FROM
    CustomerWeeklyCounts C
JOIN
    STORES S ON S.STORE_CODE = C.STORE_CODE
WHERE
    C.WEEKLY_COUNT >= 13 — Consecutive weeks of shopping
GROUP BY
    S.STORE_LOC, C.YEAR, C.QUARTER
ORDER BY
    S.STORE_LOC, C.YEAR, C.QUARTER;
```

2) Visualized via Tableau as graph titled "KPI 4". See the Tableau file.
3) Data underpins Figure 4 in the comparative analysis.

**Additional Notes:** Customer is considered habitual if they shop for at least 13 out of 16 weeks in a quarter. This shows the regularity of the customer base and a trend in the time period in each store.

## KPI 5: Rank of days of the week as per sales

**KPI Description (in words):** Ranking days of the week for each store in terms of % sales (busy days)

**KPI formula:** Count(customers), per quarter(min(purchased at)), per store

**Steps to realize KPI:**
1) Write the query below and load the extracted output as data source in Tableau

```
WITH TOT_SALES AS (
  SELECT R.STORE_CODE,
  SUM(RL.VALUE) TOT_SALES
  FROM RECEIPTS R,
  RECEIPT_LINES RL
  WHERE R.RECEIPT_ID = RL.RECEIPT_ID
  GROUP BY 1
)
SELECT S.STORE_LOC,
DATE_FORMAT(R.PURCHASED_AT, 'EEEE') AS DAY_OF_WEEK,
SUM(RL.VALUE)/T.TOT_SALES SALE_PCT,
RANK() OVER (PARTITION BY S.STORE_LOC ORDER BY SUM(RL.VALUE)/T.TOT_SALES DESC) AS
rank
FROM RECEIPTS R,
RECEIPT_LINES RL,
STORES S,
TOT_SALES T
WHERE R.RECEIPT_ID = RL.RECEIPT_ID
AND S.STORE_CODE = R.STORE_CODE
AND R.STORE_CODE = T.STORE_CODE
GROUP BY S.STORE_LOC, DATE_FORMAT(R.PURCHASED_AT, 'EEEE'), T.TOT_SALES
ORDER BY 1, 4
```

2) Visualized via Tableau as graph titled "KPI 5". See the Tableau file.
3) Data underpins Figure 5 in the comparative analysis.

**Additional Notes:** Here, each day of the week starting from Sunday to Monday is represented by the percentage of total sales occurring on that day. This is calculated for each store.

## KPI 6: Cohort Retention Analysis

**KPI Description (in words):** Cohort analysis for customer retention in each store

**KPI formula:** (Count(quarter wise customers)/Cohort Total)*100, per cohort, per quarter, per store

**Steps to realize KPI:**
1) Write the query below and load the extracted output as data source in Tableau

```
WITH COHORT_ASSIGNMENT AS (
  SELECT CUSTOMER_ID,
         DATE_TRUNC('quarter', MIN(PURCHASED_AT::DATE)) AS COHORT_QTR_DATE,
         YEAR(MIN(PURCHASED_AT::DATE)) AS COHORT_YEAR,
         STORE_CODE
```

```
  FROM RECEIPTS
  GROUP BY CUSTOMER_ID, STORE_CODE
),
COHORT_QTR_CTS AS (
  SELECT COHORT_QTR_DATE,
         EXTRACT(QUARTER FROM PURCHASED_AT::DATE) - EXTRACT(QUARTER FROM
COHORT_QTR_DATE) + 4*(YEAR(PURCHASED_AT)-COHORT_YEAR) AS RELATIVE_PERIOD,
         STORE_CODE,
         COUNT(DISTINCT CUSTOMER_ID) AS ACTIVE_CT
  FROM RECEIPTS
  JOIN COHORT_ASSIGNMENT USING (CUSTOMER_ID, STORE_CODE)
  GROUP BY COHORT_QTR_DATE, RELATIVE_PERIOD, STORE_CODE
),
COHORT_TOTALS AS (
  SELECT COHORT_QTR_DATE, STORE_CODE, COUNT(DISTINCT CUSTOMER_ID) AS COHORT_TOTAL
  FROM COHORT_ASSIGNMENT
  GROUP BY COHORT_QTR_DATE, STORE_CODE
),
COHORT_QTR_PERCENT AS (
  SELECT COHORT_QTR_DATE, STORE_CODE, RELATIVE_PERIOD, (ACTIVE_CT /
COHORT_TOTAL)*100 AS ACTIVE_PERCENT
  FROM COHORT_QTR_CTS
  JOIN COHORT_TOTALS USING (COHORT_QTR_DATE, STORE_CODE)
)
SELECT STORE_LOC, CONCAT(EXTRACT(YEAR FROM COHORT_QTR_DATE),' Q', EXTRACT(QUARTER
FROM COHORT_QTR_DATE)) AS ROW_ID,
       RELATIVE_PERIOD::STRING COL_ID,
       ACTIVE_PERCENT AS VAL
FROM COHORT_QTR_PERCENT CT,
STORES S
WHERE CT.STORE_CODE = S.STORE_CODE
UNION ALL
SELECT STORE_LOC, CONCAT(EXTRACT(YEAR FROM COHORT_QTR_DATE),' Q', EXTRACT(QUARTER
FROM COHORT_QTR_DATE)) AS ROW_ID,
       'TOTAL' AS COL_ID,
       COHORT_TOTAL AS VAL
FROM COHORT_TOTALS CT,
STORES S
WHERE CT.STORE_CODE = S.STORE_CODE
ORDER BY 1, 2, 3;
```

2) Visualized via Tableau as graph titled "KPI 6". See the Tableau file.
3) Data underpins Figure 6 in the comparative analysis.

**Additional Notes:** Customer cohort here is defined quarter-wise, so the a customer belongs to a cohort if they shopped for the first time, in that quarter. The relative period is also calculated quarter-wise. Totals represent the size of the cohort.

**KPI 7: Top 5 performing departments**

**KPI Description (in words):** Top 5 departments in each store, per year

**KPI formula:** SUM(value per department)/Total Sales), per year, per store

**Steps to realize KPI:**
1) Write the query below and load the extracted output as data source in Tableau

```
WITH
TOT_STORE_SALES AS (
  SELECT
   R.STORE_CODE,
   YEAR(R.PURCHASED_AT) AS YEAR,
   SUM(RL.VALUE) AS TOTAL_SALES
  FROM RECEIPTS R,
   RECEIPT_LINES RL
  WHERE R.RECEIPT_ID = RL.RECEIPT_ID
  GROUP BY R.STORE_CODE, YEAR(R.PURCHASED_AT)
),
RANKED_DEPT_SALES AS (
  SELECT
    S.STORE_LOC,
    YEAR(R.PURCHASED_AT) AS YEAR,
    P.DEPARTMENT_NAME,
    SUM(RL.VALUE)/T.TOTAL_SALES AS DEPT_SALES_PRT,
    ROW_NUMBER() OVER (PARTITION BY S.STORE_LOC, YEAR(R.PURCHASED_AT) ORDER BY
SUM(RL.VALUE)/T.TOTAL_SALES DESC) AS rn
  FROM
    RECEIPTS R,
    RECEIPT_LINES RL,
    PRODUCTS P,
    STORES S,
    TOT_STORE_SALES T
  WHERE R.RECEIPT_ID = RL.RECEIPT_ID
  AND RL.PRODUCT_CODE = P.PRODUCT_CODE
  AND R.STORE_CODE = S.STORE_CODE
  AND T.STORE_CODE = R.STORE_CODE
  AND YEAR(R.PURCHASED_AT) = T.YEAR
  GROUP BY
    S.STORE_LOC, YEAR(R.PURCHASED_AT), P.DEPARTMENT_NAME, T.TOTAL_SALES
)
SELECT
  STORE_LOC,
  YEAR,
  DEPARTMENT_NAME,
  DEPT_SALES_PRT*100 DEPT_SALE_PCT
FROM
  RANKED_DEPT_SALES
WHERE
  rn <= 5
ORDER BY
  STORE_LOC, YEAR, DEPARTMENT_NAME, rn;
```

2) Visualized via Tableau as graph titled "KPI 7". See the Tableau file.
3) Data underpins Figure 7 in the comparative analysis.

**Additional Notes:** Departments are represented by the proportion of sales in each store, every year. It shows only the top 5 departments in each store that contribute to the sales the most.

## Data Cleaning Steps:

1. Add a new column called store_loc to uniquely identify stores by their names and update it.

```
-- Add a new column called store_loc to the existing stores table
ALTER TABLE stores
ADD COLUMN store_loc STRING;

-- Update the new column with the concatenated values
UPDATE stores
SET store_loc = CONCAT(SUBSTRING_INDEX(address, ', ', -1), ' ', SPLIT(postcode, '
')[0]);
```

2. Delete receipt lines where quantities or product codes are null or have undefined departments

```
DELETE FROM RECEIPT_LINES
WHERE QTY IS NULL
OR PRODUCT_CODE IS NULL
OR PRODUCT_CODE IN (SELECT PRODUCT_CODE
                    FROM PRODUCTS
                    WHERE DEPARTMENT_CODE = -2)
```

3. Eliminate the anomaly found in product costs of receipt - 15512

```
UPDATE RECEIPT_LINES RL
SET VALUE = (
  SELECT AVG(VALUE/QTY)
  FROM RECEIPT_LINES
  WHERE PRODUCT_CODE = 9483
  AND RECEIPT_ID <> 15512
)
WHERE RECEIPT_ID = 15512
AND PRODUCT_CODE = 9483;

UPDATE RECEIPT_LINES RL
SET VALUE = (
  SELECT AVG(VALUE/QTY)
  FROM RECEIPT_LINES
  WHERE PRODUCT_CODE = 1988
  AND RECEIPT_ID <> 15512
)
WHERE RECEIPT_ID = 15512
AND PRODUCT_CODE = 1988;
```

## Section 2: Comparative Analysis of the KPIs

Figure 1 shows the total sales for each store neutralized by the customer count to get a fair idea of how well the stores are performing in comparison to each other. **Nottingham NG1** emerges as the top-performing store, boasting the highest average total sales per customer. It is followed by London SW1X, Birmingham B1, and London E1, in descending order. This metric considers customer count, ensuring a **balanced evaluation** of each store's efficiency in generating sales.

Analyzing the quarterly sales trends reveals a common pattern across all stores, characterized by a decrease in sales from 2020 to 2021. This decline suggests a broader trend affecting all locations during this period.

However, a notable deviation is observed in the performance of London SW1X, where the sales per customer experienced a significant dip from 94.5 in Q2 2021 to 74.4 in Q3 2021. This marked decrease suggests a unique challenge or external factor impacting the store's revenue during that specific quarter.

Interestingly, London SW1X exhibits resilience as it swiftly recovers from the downturn, returning to its earlier sales values in subsequent quarters. This recovery emphasizes the store's adaptability and ability to navigate challenges.

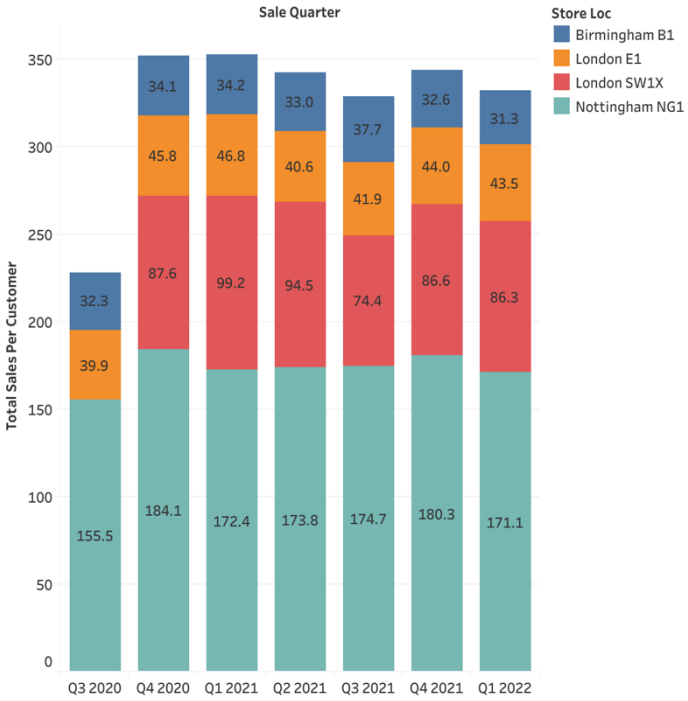


*Figure 1: Total Sales per customer at various locations*

Figure 2 sheds light on customer spending dynamics across all stores, underscoring their average Customer Lifetime Value (CLV).

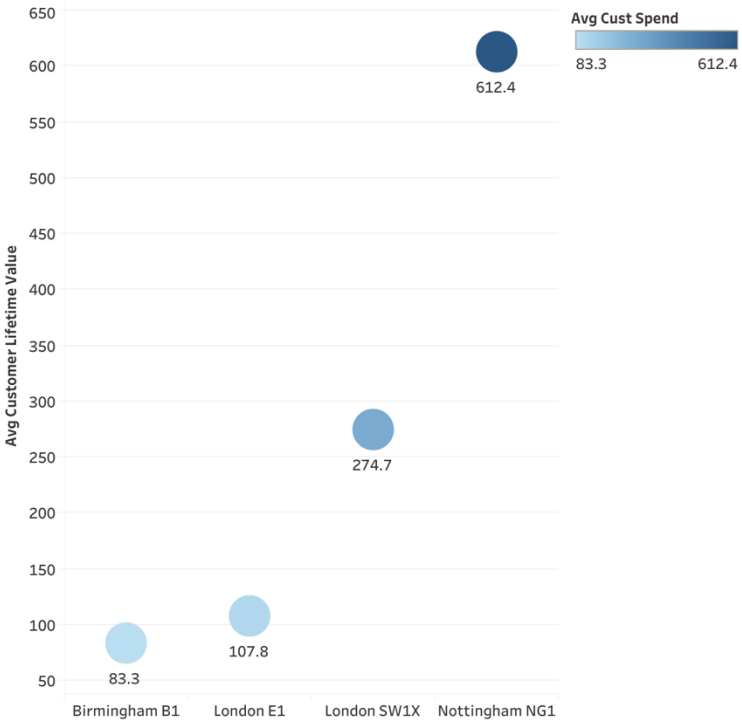


*Figure 2: Average Customer Lifetime Value(CLV) of stores*

Birmingham B1 emerges with the lowest average customer spend, translating to a comparatively modest CLV of GBP 83.34. London E1 follows with a moderate CLV of GBP 107.84, positioning it in the middle tier among the stores. However, it's **London SW1X** that notably distinguishes itself with a significantly higher average customer spend, leading to an impressive **CLV of GBP 274.69**. This store exhibits a robust potential for generating sustained revenue from its customer base.

Nottingham NG1, on the other hand, takes the top position with the highest average customer spend and an exceptional CLV of $612.39, reflecting a substantial capacity for cultivating enduring customer relationships and maximizing long-term value.

Figure 3 shows the new customer acquisition trend for various store locations. This analysis reveals a common trend of consistent decline in acquiring new customers. Despite this shared characteristic, London SW1X stands out again. The graph indicates a substantial drop in new customer acquisition for London SW1X, with the number falling from 149 in Q4 2020 to just 15 in Q1 2022. This stark decline suggests that the challenges or factors influencing customer acquisition are **more acute in London SW1X** compared to the other stores.

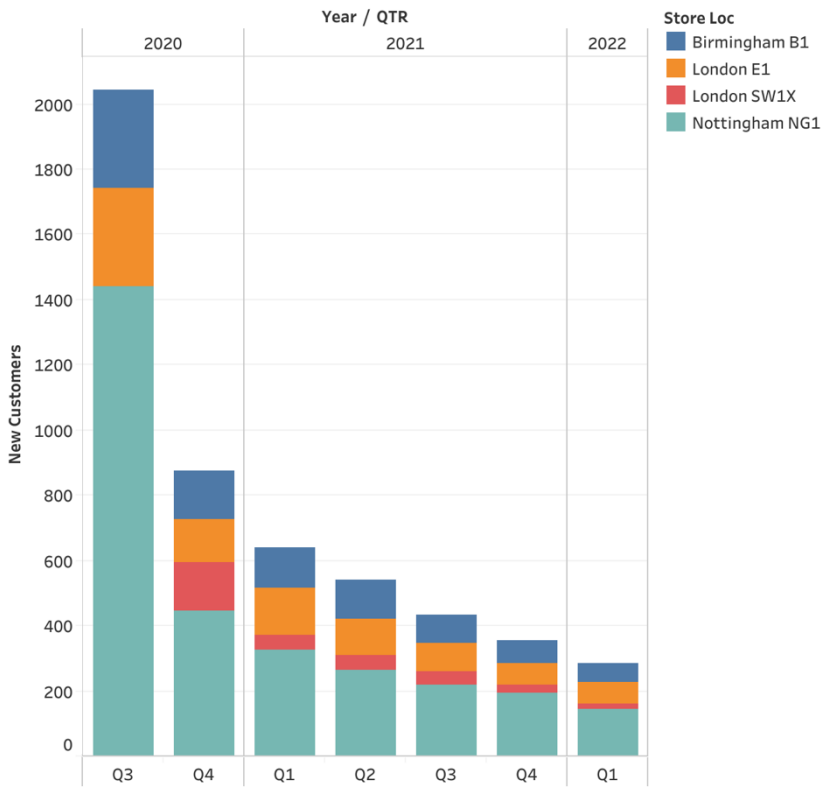


*Figure 3: Customer acquisition trend for various stores*

In Birmingham B1, a substantial surge of 300 new customers in Q3 2020 was followed by a notable decrease to 149 in Q4 2020. The trend continued with a steady decline throughout 2021, reaching 57 new customers in Q1 2022. The trend was similar for London E1 and Nottingham NG1 as well.

The significance of the declining trend in London SW1X implies a pressing need for targeted marketing strategies and investments specifically tailored to this store. Implementing strategies aimed at **revitalizing new customer acquisition in London SW1X** will be key in restoring and enhancing the store's ability to attract a fresh customer base, ensuring its continued success in a competitive retail business.

Figure 4 shows the habitual customer behavior across store locations. Nottingham NG1 emerges as the leader with consistently high numbers, showcasing a robust count of habitual customers, peaking at 168 in Q4 2021. Following Nottingham NG1, **London SW1X exhibits a commendable performance**, with its fluctuating pattern culminating in a peak of 16 habitual customers in Q2 2021. London E1 follows suit with a gradual decline from its peak of 12 habitual customers in Q4 2020, and Birmingham B1, while demonstrating growth in habitual customers in 2021, concludes with a slight decline to 2 habitual customers in Q1 2022.

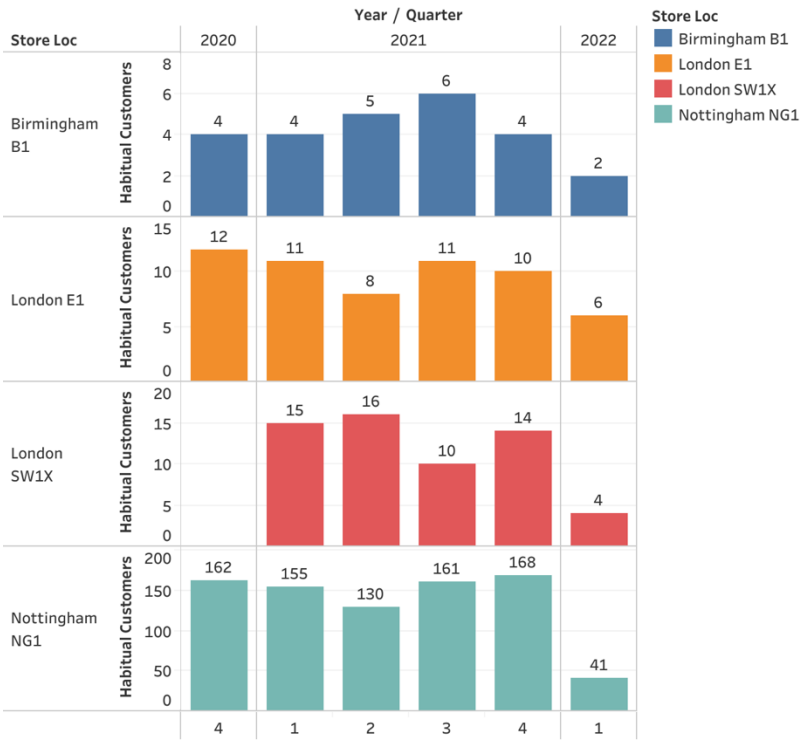


*Figure 4: Habitual Customers in each store*

A strategic consideration arises when examining London SW1X, as the store's potential for habitual customer engagement is evident despite the fluctuating trend. With a targeted marketing investment, London SW1X can not only stabilize but significantly increase its habitual customer count.

Figure 5 shows the proportion of sales for each day of the week across different store locations. Notably, Nottingham NG1 stands out with a distinct sales distribution, where Saturday dominates as the highest sales day, accounting for a substantial 18.8%, followed closely by Friday at 17.6%. This unique sales pattern suggests a notable preference among customers in Nottingham NG1 for **weekend shopping.**


*Figure 5: Proportion of sales per day of the week*

For London SW1X, Friday emerges as the top sales day, representing 17.9% of total sales. Saturday follows closely behind, contributing 14.3%. The prominence of Friday and Saturday in **London SW1X** suggests a potential concentration of customer activity during the end of the week, like the top performer Nottingham NG1. This indicates an opportunity to optimize marketing strategies to capitalize on these high-traffic days.

Contrastingly, Birmingham B1 and London E1 exhibit a more evenly distributed sales pattern throughout the week, with each day contributing between 13.5% and 15.6%.

Figure 6 shows cohort retention analysis presents a comprehensive overview of customer retention percentages for each quarter across different store locations. Birmingham B1 and London E1 exhibit consistent declines, while London SW1X, despite fluctuations, shows decreasing retention percentages. Nottingham NG1 experiences moderate declines. Considering the affluent market in **London SW1X**, strategic investments to address retention issues could be more beneficial, aligning with the store's unique demographic advantages compared to other locations.

| | | Relative Period | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Store Location | Quarters | 0 | 1 | 2 | 3 | 4 | 5 | 6 | TOTAL |
| **Birmingham B1** | 2020 Q3 | 100 | 58 | 53 | 50 | 52 | 52 | 43 | 300 |
| | 2020 Q4 | 100 | 32 | 30 | 28 | 26 | 21 | | 149 |
| | 2021 Q1 | 100 | 28 | 20 | 23 | 20 | | | 123 |
| | 2021 Q2 | 100 | 17 | 17 | 16 | | | | 119 |
| | 2021 Q3 | 100 | 14 | 13 | | | | | 87 |
| | 2021 Q4 | 100 | 18 | | | | | | 71 |
| | 2022 Q1 | 100 | | | | | | | 57 |
| **London E1** | 2020 Q3 | 100 | 58 | 50 | 48 | 46 | 45 | 44 | 301 |
| | 2020 Q4 | 100 | 36 | 34 | 35 | 28 | 32 | | 133 |
| | 2021 Q1 | 100 | 33 | 28 | 21 | 18 | | | 145 |
| | 2021 Q2 | 100 | 27 | 14 | 16 | | | | 113 |
| | 2021 Q3 | 100 | 31 | 22 | | | | | 89 |
| | 2021 Q4 | 100 | 32 | | | | | | 69 |
| | 2022 Q1 | 100 | | | | | | | 68 |
| **London SW1X** | 2020 Q4 | 100 | 80 | 77 | 69 | 60 | 61 | | 149 |
| | 2021 Q1 | 100 | 43 | 36 | 48 | 43 | | | 42 |
| | 2021 Q2 | 100 | 38 | 36 | 19 | | | | 42 |
| | 2021 Q3 | 100 | 33 | 28 | | | | | 40 |
| | 2021 Q4 | 100 | 27 | | | | | | 22 |
| | 2022 Q1 | 100 | | | | | | | 15 |
| **Nottingham NG1** | 2020 Q3 | 100 | 78 | 74 | 72 | 68 | 68 | 64 | 1,443 |
| | 2020 Q4 | 100 | 42 | 40 | 32 | 38 | 31 | | 445 |
| | 2021 Q1 | 100 | 32 | 31 | 29 | 25 | | | 328 |
| | 2021 Q2 | 100 | 29 | 26 | 23 | | | | 266 |
| | 2021 Q3 | 100 | 27 | 23 | | | | | 219 |
| | 2021 Q4 | 100 | 32 | | | | | | 195 |

*Figure 6: Cohort Retention Analysis of each store*

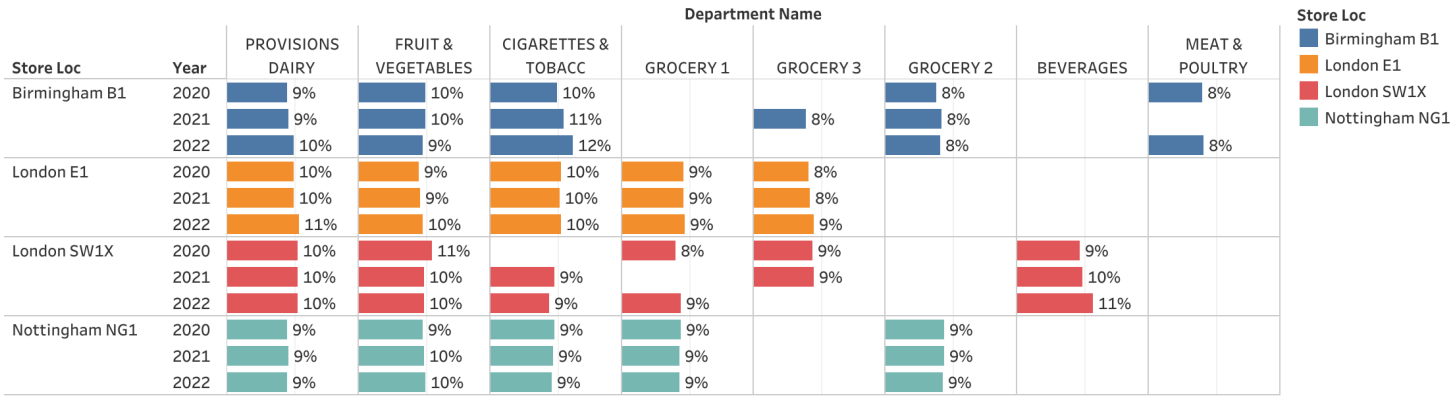


*Figure 7: Top 5 Departments in each of the stores and their share in sales*

Figure 7 shows the top-performing departments across various stores and years. Nottingham NG1 demonstrates a balanced performance across "Cigarettes & Tobacco," "Fruit & Vegetables," and "Provisions Dairy," maintaining steady sales percentages. Most departments in other stores have demonstrated consistent performance each year, with some minor fluctuations. Notably, the **Beverages department in London SW1X** experienced a remarkable shift, accounting for 9% of total sales in 2020 and surging to 11% in 2022. This notable growth suggests that investing in the Beverages department in London SW1X could yield greater efficiency compared to other stores, leveraging the distinctive demographic and local market trends of the area.

However, it's crucial to acknowledge the inconsistency observed in London SW1X, particularly in Grocery-related departments. While **Grocery 1 and Grocery 2** have proven successful in other stores, London SW1X exhibits varying performance in these areas. Recognizing the success of these departments elsewhere, targeted investments in Grocery 1 and Grocery 2 for London SW1X could be highly effective, aligning with established customer preferences and potentially enhancing overall store performance.

In conclusion, the comprehensive analysis of various key performance indicators strongly suggests that directing marketing investments towards **London SW1X would be the most effective strategy**. Despite facing challenges in new customer acquisition and experiencing fluctuations in certain departments, London SW1X exhibits unique potential for growth. The store's consistently high average customer lifetime value and a notable increase in the "Beverages" department make it a promising candidate for strategic investments. By addressing specific issues, such as the decline in new customer acquisition and optimizing departmental performance, a targeted marketing approach in London SW1X could yield substantial returns, making it the optimal choice for maximizing overall store performance.