
# Data At Scale

# Coursework

# 2020-21

**Submitted To –**

**Dr. Gavin Smith**

**Submitted By-**

**Jignesh Manocha**

**ID- 20249953**

# Section 1

# The KPIs

## Store KPIs

**KPI Description (in words):** Total Revenue / Store

**KPI formula:** Sum of Value, Group by store

**Steps to realize KPI:**

1) SELECT SUM(value) AS total_revenue
   store_code
   FROM w19.receipts
   JOIN w19.receipt_lines
   USING (receipt_id)
   WHERE value != 9999999
   GROUP BY 2
   ORDER BY 2;

2) Visualized via Tableau as graph titled "Total Revenue". See the Tableau file.

**Additional Notes:** Illegal Values (9999999) has not been taken into consideration.

**KPI Description (in words):** Average Transaction Value / Store

**KPI formula:** Sum of Value / Count(distinct receipt_id), Group by store_code

**Steps to realize KPI:**

1) CREATE VIEW active_customers AS
   SELECT SUM(value) / COUNT(DISTINCT receipt_id) AS avg_transaction_value,
   store_code
   FROM w19.receipts
   JOIN w19.receipt_lines
   USING(receipt_id)
   WHERE value != 9999999
   GROUP BY store_code;

2) Visualized via Tableau as graph titled "Average Transaction Value". See the Tableau file.

**Additional Notes:** Illegal Values (9999999) has not been taken into consideration.

**KPI Description (in words):** Number of Departments / Store

**KPI formula:** Count of distinct department's code, group by Store code

**Steps to realize KPI:**

1) CREATE VIEW departments_per_store AS
   SELECT COUNT(DISTINCT department_code) AS ct, store_code
   FROM w19.receipts
   JOIN w19.receipt_lines
   USING (receipt_id)
   JOIN w19.products
   USING (product_code)
   GROUP BY 2;

2) Visualized via Tableau as graph titled "Departments Per Store". See the Tableau file.

**Additional Notes:** None

**KPI Description (in words):** Count of High Sale Weekdays / Store / Month

**KPI formula:** Count of weekdays where sum of value is greater than 500, Group by weekdays, store code and month

**Steps to realize KPI:**

1) CREATE VIEW high_sale_weekdays AS
   SELECT month, COUNT(*) as ct, store_code
   FROM (
      SELECT EXTRACT(isodow FROM purchased_at) AS dow,
      to_char(DATE_TRUNC('month',purchased_at),'YYYY-MM') AS month,
      store_code
      FROM w19.receipts
      JOIN w19.receipt_lines
      USING (receipt_id)
      WHERE EXTRACT(isodow FROM purchased_at) < 6
      AND value != 9999999
      GROUP BY 1, 2,3
      HAVING SUM(value) > 500
   )x
   GROUP BY month, store_code
   ORDER BY store_code;

2) Visualized via Tableau as graph titled "High Sale Weekdays". See the Tableau file

**Additional Notes:** It has been assumed that high sale days are days when as store sells over 500£ of goods. High sale days are days when a store sells over £100 of goods.

## Customer KPIs

**KPI Description (in words):** Active Customers/Month/Store

**KPI formula:** Active Customers – count unique customers
Per Month / Per Store – Group by month and store

**Steps to realize KPI:**

1) CREATE VIEW active_customers AS
SELECT COUNT(DISTINCT customer_id)AS active_ct,
to_char(DATE_TRUNC('month',purchased_at),'YYYY-MM') AS month,
store_code
FROM w19.receipts
GROUP BY 2,3
ORDER BY store_code;

2) Visualized via Tableau as graph titled "Active Customers Per Month". See the Tableau file.

**Additional Notes:** None.

**KPI Description (in words):** Loyal Customers/Month/Store

**KPI formula:** Count( distinct purchased_at) > = 3, Group by customer_id, store code and month.

**Steps to realize KPI:**

1) CREATE VIEW loyal_customers AS
SELECT COUNT(*) AS loyal_customer_count, month, store_code
FROM(
SELECT customer_id,
to_char(DATE_TRUNC('month',purchased_at::DATE),'YYYY-MM') AS month,
store_code
FROM w19.receipts
GROUP BY 1,2,3
HAVING COUNT(DISTINCT(purchased_at)) >= 3
)x
GROUP BY month, store_code
ORDER BY store_code;
2) Visualized via Tableau as graph titled "Loyal Customers Per month". See the Tableau file.

**Additional Notes:** Customers are assumed to be loyal who made purchase more than thrice a month.

**KPI Description (in words):** New Customers/Month/Store

**KPI formula:** Count (MIN(purchased_at)), per month, per store, GROUP BY customer_id and store_code

**Steps to realize KPI:**

1) CREATE VIEW new_customers AS
   SELECT COUNT(*) as ct,first_mth, store_code
   FROM(
      SELECT customer_id,
      to_char(DATE_TRUNC('month', MIN(purchased_at)),'YYYY-MM') AS first_mth,
      store_code
      FROM w19.receipts
      GROUP BY 1,3
   )x
   GROUP BY first_mth, store_code
   ORDER BY store_code;

2) Visualized via Tableau as graph titled "New Customers Per Month". See the Tableau file.

**Additional Notes:** First date of customer shopping with store is assumed as new customer.

**KPI Description (in words):** Repeat Customer / Store / Week

**KPI formula:** Count(distinct purchased_at) > 1, Group by store and week

**Steps to realize KPI:**

1) CREATE VIEW repeat_customers AS
   SELECT count(*) AS ct, week, store_code
   FROM(
      SELECT customer_id,
      DATE_TRUNC('week', purchased_at)::DATE as week,
      store_code
      FROM w19.receipts
      JOIN w19.receipt_lines
      USING (receipt_id)
      GROUP BY 1,2,3
      HAVING COUNT(DISTINCT purchased_at) > 1
   )x
   GROUP BY week, store_code
   ORDER BY store_code;
2) Visualized via Tableau as graph titled "Repeat Customers Per Week". See the Tableau file

**Additional Notes:** Assumption- Customers coming in two different days in a week are considered as repeating customers.

**KPI Description (in words):** Cohort Analysis Nottingham Store (0)

**KPI formula:** Count active customers per relative period, Calculate percent, Count number of customers per cohort.

Steps to realize KPI:

1) 
```
CREATE TABLE cohort_analysis_nottingham_store AS
WITH cohort_assignment AS (
        SELECT customer_id, date_trunc('month', MIN(purchased_at))::DATE AS cohort_date
        FROM w19.receipts
        WHERE store_code = 0
        GROUP BY customer_id
),
cohort_mth_counts AS (
        SELECT cohort_date, EXTRACT(month from age(purchased_at, cohort_date)) +
        12 * EXTRACT(year from age(purchased_at, cohort_date)) as relative_period,
        COUNT(DISTINCT(customer_id)) AS active_count
        FROM w19.receipts
        JOIN cohort_assignment
        USING (customer_id)
        WHERE purchased_at >= cohort_date
        GROUP BY cohort_date, relative_period
),
cohort_totals AS (
        SELECT cohort_date, COUNT(DISTINCT customer_id) AS cohort_total
        FROM cohort_assignment
        GROUP BY cohort_date
),
cohort_mth_percent AS (
        SELECT cohort_date, relative_period, active_count :: NUMERIC /
        cohort_total AS active_percent
        FROM cohort_mth_counts
        JOIN cohort_totals
        USING (cohort_date)
)
SELECT cohort_date AS row_id, relative_period :: TEXT AS col_id, active_percent AS value
FROM cohort_mth_percent
UNION ALL
SELECT cohort_date AS row_id, 'total'::TEXT AS col_id, cohort_total AS value
FROM cohort_totals;
```

2) Visualized via Tableau as graph titled "Cohort_Analysis_Nottingham_0". See the Tableau file.

**Additional Notes:** None

**KPI Description (in words):** Cohort Analysis Birmingham Store (1)

**KPI formula:** Count active customers per relative period, Calculate percent, Count number of customers per cohort.

Steps to realize KPI:

1) CREATE TABLE cohort_analysis _birmingham_store AS
WITH cohort_assignment AS (
SELECT customer_id, date_trunc('month', MIN(purchased_at))::DATE AS cohort_date
FROM w19.receipts
WHERE store_code = 1
GROUP BY customer_id
),
cohort_mth_counts AS (
SELECT cohort_date, EXTRACT(month from age(purchased_at, cohort_date)) +
12 * EXTRACT(year from age(purchased_at, cohort_date)) as relative_period,
COUNT(DISTINCT(customer_id)) AS active_count
FROM w19.receipts
JOIN cohort_assignment
USING (customer_id)
WHERE purchased_at >= cohort_date
GROUP BY cohort_date, relative_period
),
cohort_totals AS (
SELECT cohort_date, COUNT(DISTINCT customer_id) AS cohort_total
FROM cohort_assignment
GROUP BY cohort_date
),
cohort_mth_percent AS (
SELECT cohort_date, relative_period, active_count :: NUMERIC /
cohort_total AS active_percent
FROM cohort_mth_counts
JOIN cohort_totals
USING (cohort_date)
)
SELECT cohort_date AS row_id, relative_period :: TEXT AS col_id, active_percent AS value
FROM cohort_mth_percent
UNION ALL
SELECT cohort_date AS row_id, 'total'::TEXT AS col_id, cohort_total AS value
FROM cohort_totals;

2) Visualized via Tableau as graph titled " Cohort_Analysis_Birmingham_1". See the Tableau file.

**Additional Notes:** None

**KPI Description (in words):** Cohort Analysis London Store 2

**KPI formula:** Count active customers per relative period, Calculate percent, Count number of customers per cohort.

Steps to realize KPI:

1) 
```
CREATE TABLE cohort_analysis _london_store _2 AS
WITH cohort_assignment AS (
        SELECT customer_id, date_trunc('month', MIN(purchased_at))::DATE AS
cohort_date
        FROM w19.receipts
        WHERE store_code = 2
        GROUP BY customer_id
),
cohort_mth_counts AS (
        SELECT cohort_date, EXTRACT(month from age(purchased_at, cohort_date)) +
        12 * EXTRACT(year from age(purchased_at, cohort_date)) as relative_period,
        COUNT(DISTINCT(customer_id)) AS active_count
        FROM w19.receipts
        JOIN cohort_assignment
        USING (customer_id)
        WHERE purchased_at >= cohort_date
        GROUP BY cohort_date, relative_period
),
cohort_totals AS (
        SELECT cohort_date, COUNT(DISTINCT customer_id) AS cohort_total
        FROM cohort_assignment
        GROUP BY cohort_date
),
cohort_mth_percent AS (
        SELECT cohort_date, relative_period, active_count :: NUMERIC /
        cohort_total AS active_percent
        FROM cohort_mth_counts
        JOIN cohort_totals
        USING (cohort_date)
)
SELECT cohort_date AS row_id, relative_period :: TEXT AS col_id, active_percent AS value
FROM cohort_mth_percent
UNION ALL
SELECT cohort_date AS row_id, 'total'::TEXT AS col_id, cohort_total AS value
FROM cohort_totals;
```

2) Visualized via Tableau as graph titled " Cohort_Analysis_London_2". See the Tableau file.

**Additional Notes:** None

**KPI Description (in words):** Cohort Analysis London Store 3

**KPI formula:** Count active customers per relative period, Calculate percent, Count number of customers per cohort.

Steps to realize KPI:

1) CREATE TABLE cohort_analysis _london_store _3 AS
WITH cohort_assignment AS (
SELECT customer_id, date_trunc('month', MIN(purchased_at))::DATE AS cohort_date
FROM w19.receipts
WHERE store_code = 3
GROUP BY customer_id
),
cohort_mth_counts AS (
SELECT cohort_date, EXTRACT(month from age(purchased_at, cohort_date)) +
12 * EXTRACT(year from age(purchased_at, cohort_date)) as relative_period,
COUNT(DISTINCT(customer_id)) AS active_count
FROM w19.receipts
JOIN cohort_assignment
USING (customer_id)
WHERE purchased_at >= cohort_date
GROUP BY cohort_date, relative_period
),
cohort_totals AS (
SELECT cohort_date, COUNT(DISTINCT customer_id) AS cohort_total
FROM cohort_assignment
GROUP BY cohort_date
),
cohort_mth_percent AS (
SELECT cohort_date, relative_period, active_count :: NUMERIC /
cohort_total AS active_percent
FROM cohort_mth_counts
JOIN cohort_totals
USING (cohort_date)
)
SELECT cohort_date AS row_id, relative_period :: TEXT AS col_id, active_percent AS value
FROM cohort_mth_percent
UNION ALL
SELECT cohort_date AS row_id, 'total'::TEXT AS col_id, cohort_total AS value
FROM cohort_totals;

2) Visualized via Tableau as graph titled " Cohort_Analysis_London_3". See the Tableau file.

**Additional Notes:** None

# Section 2

## Executive Summary

This report provides an analysis of regional performance of FoodCorp's stores operating in Nottingham, Birmingham, and London using metric measures in order to determine the potential store to plan a marketing strategy and distribution tactics. Method of evaluation include key performance indicators (KPIs) of growth, engagement, and retention such as active, repeat customers, average transaction value and cohort analysis.

Investigation through KPIs reveals that London Store 2 is the largest and performing best in every perspective such as customer base, trends, and sales followed by Store 3, Birmingham, and Nottingham store. Store 2 contributes around 83.2% in total revenue of the company, has customer counts of 2888, high transactional value. However, this is not enough to complete the argument. Considering company's goal of selecting the store which provides the optimum profit with respect to its size, results of data analysed indicates that store 3 of London has the best potential for growth as it performed well in every measure after store 2. Store 3 contributes 8.22% of total revenue, has maximum average transaction value with 11.31 per transaction, 36 departments, 99 high sales weekdays. Additionally, customer count is also highest in this store with 1275 customers, has highest number of loyal and new customers. Further examination reveals that customer retention is not so good in this store as it retains only 50% customers initially on an average, which depicts that this store is capable to attract new customers but lack in retaining them.

It is significant to consider customer retention, as it has the potential to drop the overall profit, sales of the company. Thus, company needs to make strategies and plan to tackle this problem to yield maximum return on investment and make the store 3 profitable. It is recommended that -

- FoodCorps should conduct the marketing campaign in the store 3 of London as it completes the argument and is capable enough to contribute much to the company by attracting more customers and retaining existing ones.
- The company should make Public Relations strategy in order to improve customer retention and loyalty.
- The company should implement promotional mix tools to attract customers and conduct analysis on marketing mix tools.

# Section 3

# Comparative Analysis

## Store size Based

### Total Revenue per store

London Store 2 has the maximum total revenue with 83.2% share among all stores followed by Store 3. Birmingham and Store 3 with 5.20% and 8.22% contribution, respectively. Nottingham Store shares the lowest revenue to FoodCorps with only 3.34%. Over the period, in terms of money store 2,3,1,0 earned the total revenue of 1512901.69£,149476.60£, 94483.62£ and 60637.96£ respectively, in which there is a huge difference between highest store and second-best store because of the high value customers and some illegal values in store 2.

**Note** – Some illegal values have been founded in the data. Thus, for better results it has been not taken into consideration.

### Average Transaction Value

Average Transactional Value help to plan the marketing offers which is the major concern for FoodCorps. In analysis, the proportional value with respect to the transaction made is also maximized by store 2 with 18.06. In store 3, consumers are spending 11.37 on an average, followed by Birmingham with 11.13 spending. Higher the average receipt value higher will be the revenue earned per customer. Since store 2 includes many illegal values such as 9999999 which might skew the final ATV, therefore, store 3 has the next best potential to earn more amount per transaction.

### Department Per Store

To evaluate the size of the stores, it is essential to know the number of departments in the store. At the time of analysis, it has been found that Store 2 has 40 departments whereas store 0,1,3 have 32,35 and 36, respectively. Since number of departments is highest in store 2 and second most in store 3 followed by store 1, it might be concluded that store 2 is biggest among all stores.

### High Sale Weekdays

This KPI will help to analyse the number of weekdays with high number of footfalls and high value customers with respect to the store's size.

There is total count of 103 high sale weekdays in store 2, which is highest in all stores and store 3 have second most weekdays i.e., 99 weekdays followed by store 1 and 0 with 85 and 29 weekdays in its total count, respectively.

With respect to the size of the store, store 3 would be considered best in high-sale weekdays.

In order to examine the total customers over the time, it has been found that Store 2 in London has the maximum customer base of 2888 customers followed by 1275 customers in store 3. The store 1 and 0 has the lowest customer base in all stores with 753 and 659, respectively.

Since the total revenue, customer base and number of departments are high in Store 2, it can be concluded that it is the largest store in terms of size, and store 3 in London follows it with second best mentioned parameters which proves its potential. Birmingham and Nottingham Store are the lowest contributors to the FoodCorps with respect to these parameters.

NOTE – To analyse the data effectively, illegal value (9999999) has been omitted and not taken into consideration.

## Customer Based

### Active customers Per Month

This KPI reveals the growth overtime in terms of customer base of the stores. This performance measure has been chosen to evaluate the number of unique footfalls per month into the store. This also helps to analyse the potential customers, their behaviour and to assess marketing strategies effectively.

Analysis – Active customers are highest in London Store 2 with around 850 customers per month on an average. Store 3 has the next best average of around 250 customers per month over time and noted the rising trend in 2$^{nd}$ quarter when active customers in store 2 were declining.

### Repeat Customers Per Month

This metric measure indicates the customer's engagement in the store. This KPI has been considered to determine the value, which store is delivering to its customers. It has been assumed that any customers returning more than once in a week falls in this category.

Analysis-

Since the London Store 2 is relatively larger in size, the repeating count is directly proportional, Store 3 noted the rising trend of repeat customers from week 9 to 16 and from week 25 to 29. In a meantime, repeating customers in Birmingham Store were increasing on a fast pace than store 2 and 3 which illustrates that they are lagging in retrieving customers.

### New Customers

New customers are crucial to a company's growth, as it shows how much customers company is pulling. This KPI has been taken to examine the capability of all stores to gain new customers. The higher the ability to attract new customers, the higher will be the probability of growth of company in terms of revenue, sales.

Analysis – Store 2 attract the most customers relative to its size following with store 3. Store 2 and 3 performed best in first two quarters, then losing attract new customers over the time. Birmingham store were not in the picture initially but after entry into the market it showed the rising trend in starting which dipped off later. The trend of Store 0 illustrates that it has attracted very few customers over time, thus it has very less potential to gain new customers.

**Loyal Customers Per month**

Customer's loyalty is also the key metric to measure success of the company as it reveals the customer's value towards company. This KPI has been chosen to determine the potential customers of all stores and plan marketing scheme accordingly. Customers are assumed to be loyal when they come at least thrice a month in store.

Analysis- Store 2 seems to be the best store with total count of 9146 loyal customers. Store 3 achieves 2nd best performance with the customer loyalty of 1266, which reveals its potential to retrieve customers after store 2. Over the period, Store 0 and 1 are not good enough to retrieve the customers as they are retaining only 678 and 943 customers in total counts, respectively.

**Cohort Analysis**

Cohort Analysis has been chosen to examine the customer lifecycle and performance of the store to retain the customers. This metric measure helps to investigate the customer loyalty and their behaviours towards stores. Based on the results, the customer retention are analysed below-

**London Store_2**

- Customer Retention is best for this store comparative to all other stores. The retention rate for store 2 performed outstanding in the month of March 2019 and optimum in April 2019.
- Store were retrieving on an average of 75% customers in March and April of 2019 with stable average in every relative period and which noted the subsequent declining trend in later periods.
- After June 2019 it can be noted that there is static trend of customer retention with an average of 0.25, in which low value customers are may be the reason.
- During the last cohort and month retention rate was at its minimum which depicts that Store were started losing its customers in later months where lack of marketing offers might be the reason.

**London Store_3**

- Store 3 has been not performed very well in retrieving customers compare to store 2.
- London store 3 is retaining 50% customers on an average during first two months in quarter 2,2019 which eventually started declining in the later cohort months.
- In March,2019, out of 87 customers 70% customers were retained on relative day 1 and on an average 50% were retained over customer lifetime.
- On day 1, there is stable trend of customer retention after initial two months since store has been introduced. There is decline in retention in June and October 2020 on an average to 10%.

**Birmingham Store_1**

- Store 1 is slightly better in retaining customers than store 3. This store is retrieving on an average of 60% and 50% in the month of June and July 2019, respectively.
- There is stable trend on day 1 over the period except in May 2020 which has the lowest mean.
- The total customers increase in the store initially but after some time, the store unable to retrieve many customers and hence noted the imbalance trend of customer retention in total.
- December 2019 and February 2020 noted the lowest customer retention rate with 10% average customers, this might be due to the external factors like season, policies etc.

**Nottingham Store_0**

- Store 0 performed 2$^{nd}$ best in retaining the customer after store 2 in London.
- This store has an average of retrieving around 70% consumers in the month of March,2019 and 40% mean in April,2019 which are the top performing months of this store in terms of customer retention.
- However, this store has not performed well over the period as it has noted very unstable trend in retaining customers. In January 2020, this store has retained the lowest customers in the cycle with an average of 4% customers.
- There has been drastic declining trend after quarter 2 in 2019, it can be inferred that the customers who came in starting did not engage with the store again.