Final Report

03.31.2023

Khushi Saini

Jenna Lusk

Rose Elder

Jennifer Merry

Background

Our project is focused on Target, one of the largest retail corporations in the United States. As consumers of the company ourselves, we are interested in exploring the rewards realm and gaining a deeper understanding of the benefits that both customers and employees can receive through discounts, referrals, and coupons. Specifically, we are interested in understanding the factors that contribute to markdowns, and how these factors can be optimized to benefit both the company and its customers. Through our analysis of the available dataset, we aim to shed light on the rewards programs offered by Target and their impact on consumer interactions with the brand. We will investigate the relationship between discounts and referrals and consumer spending, as well as how these rewards programs can drive customer loyalty and retention. Our ultimate goal is to provide valuable insights into the rewards landscape at Target, helping both the company and its customers to better understand the value of these programs and how they can be used to create mutually beneficial relationships.

Database Description

Employee

EmployeeID	INT(9)
EmpFirstName	VARCHAR(20)
EmpLastName	VARCHAR(30)
EmpPhoneNumber	VARCHAR(12)
EmpEmail	VARCHAR(100)
JobTitle	VARCHAR(20)
Pay	DECIMAL(11,2)
StartDate	VARCHAR(8)

Customer

CustomerID (PK)	INT(9)
RewardsID (FK)	INT(9) NULL
CustFirstName	VARCHAR(20)
CustLastName	VARCHAR(30)
CustPhoneNum	VARCHAR(12)
CustEmail	VARCHAR(100)
Custbirthday	VARCHAR(10)

Rewards

RewardsID (PK)	INT(9)
RewardStartDate	VARCHAR(512)
Points	INT
RewardsLevel (bronze,silver,gold)	VARCHAR(512)

Purchase

PurchaseID (PK)	INT(9)
EmployeeID (FK)	INT(9)
RewardsUsed	VARCHAR(512)
RewardsID	INT(9) NULL
TotalCost	DECIMAL(11,2)
PointsEarned	INT(2) (can be null)
CouponUsed	VARCHAR(512)
PointsUsed	INT(2) NULL
PurchaseDate	VARCHAR(512)
Zip	INT

Products

ProductID (PK)	INT(9)
ProductDesc	VARCHAR(125)
ProductPoint	INT(1)
Price	DECIMAL(11,2)
Stock	INT(3)
FloorLocation(where its stocked)	VARCHAR(50)

Purchase_Products

ProductID (PK)	INT(9)
PurchaseID (PK)	INT(9)
Quantity	INT(2)

Solutions

What are the purchasing trends amongst different age groups?

As per comparing the trend differences, the Function information from this class will be knowledgeable in order to create procedures or functions to get ahold of all the trends. It will require taking the ProductID and ProductDesc to check and see which products are the most purchased and then through the utilization of cursors or even control structures, it will assist in the gathering of data. This is useful to know because it allows for the understanding of trends and supply chain as well as operations. This would allow for obtaining profit and allow Target to make further efforts to break even every year.

Code:

```
SELECT AgeGroup, COUNT(*) as PurchaseCount
FROM (
   SELECT
      CASE
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) BETWEEN 18 AND 24 THEN '18-24'
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) BETWEEN 25 AND 34 THEN '25-34'
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) BETWEEN 35 AND 44 THEN '35-44'
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) BETWEEN 45 AND 54 THEN '45-54'
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) BETWEEN 55 AND 64 THEN '55-64'
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) >= 65 THEN '65+'
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) <= 17 THEN 'Under 18'
         ELSE 'N/A'
      END AS AgeGroup
   FROM Customer1
```

```
   UNION ALL
   SELECT
      CASE
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) BETWEEN 18 AND 24 THEN '18-24'
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) BETWEEN 25 AND 34 THEN '25-34'
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) BETWEEN 35 AND 44 THEN '35-44'
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) BETWEEN 45 AND 54 THEN '45-54'
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) BETWEEN 55 AND 64 THEN '55-64'
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) >= 65 THEN '65+'
         WHEN EXTRACT(YEAR FROM SYSDATE) - (EXTRACT(YEAR FROM
TO_DATE(CustBirthday, 'MM/DD/YY')+1900)) <= 17 THEN 'Under 18'
         ELSE 'N/A'
      END AS AgeGroup
   FROM Purchase
   JOIN Rewards ON Purchase.RewardsID = Rewards.RewardsID
   JOIN Customer1 ON Rewards.RewardsID = Customer1.RewardsID
)
GROUP BY AgeGroup;
```

Output:

```
AGEGROUP PURCHASECOUNT
-------- -------------
Under 18          632
18-24               5
```

How many purchases are there near Purdue?

There are two zipcodes that have a Target by Purdue, 47907 and 47906. Knowing the number of purchases with a zipcode because it is useful to track purchases in a college town and to see if the Targets in that location are being successful, so that Target can begin adding new Targets on different college campuses.

Code:

```
SELECT ZIPCODE, COUNT(ZIPCODE) AS PURCHASE_COUNT
FROM Purchase
WHERE ZIPCODE = '47907' OR ZIPCODE = '47906'
GROUP BY ZIPCODE;
```

Output:

ZIPCODE	PURCHASE_COUNT
47906	10
47907	11

What is the top 10% of locations in terms of sales?

To begin with, obtaining the top 10% of locations in terms of sales it would permit for understanding which locations are doing the best. To do so, we need to focus on the SQL Tuning aspect, along with functions and exceptions. This would help create the proper structure of noting down the top Zip's and Price which would allow checking to see which factors cause high profit and can be implemented in the other Target locations.

Code:

```
SELECT *
FROM (
   SELECT
      FloorLocation,
      SUM(Price * Quantity) AS TotalSales
   FROM
      Purchase_Products pp
```

```
    INNER JOIN Product p ON pp.ProductID = p.ProductID
  GROUP BY
    FloorLocation
  HAVING
    SUM(Price * Quantity) >= (
      SELECT
        PERCENTILE_CONT(0.9) WITHIN GROUP (ORDER BY SUM(Price * Quantity) DESC)
      FROM
        Purchase_Products pp
        INNER JOIN Product p ON pp.ProductID = p.ProductID
      GROUP BY
        FloorLocation
    )
  ORDER BY
    TotalSales DESC
  )
WHERE
  ROWNUM <= (
    SELECT
      CEIL(COUNT(*) * 0.1)
    FROM
      (
        SELECT DISTINCT
          FloorLocation
        FROM
          Purchase_Products pp
          INNER JOIN Product p ON pp.ProductID = p.ProductID
      )
  );
```

Output:

FLOORLOCATION	TOTALSALES
Donec est	6640.88
dolor dapibus	5752.24
Nullam nisl. Maecenas malesuada	4617.45
Aliquam vulputate ullamcorper magna. Sed eu eros.	4371.72
Curabitur sed tortor. Integer aliquam	4283.52
auctor ullamcorper, nisl arcu	4064.86
In at pede.	3829.68
at pede. Cras vulputate velit eu sem.	3747.84
felis.	3735.92
eu nibh vulputate mauris sagittis	3699.54
mattis. Cras eget nisi dictum augue malesuada male	3301.51

FLOORLOCATION	TOTALSALES
ipsum primis in faucibus orci luctus et ultrices p	3241.92
lectus pede, ultrices a, auctor non, feugiat	3203.74
mollis. Duis sit amet diam	3101.1
Phasellus nulla. Integer vulputate, risus	3095.07
Suspendisse non	3074.87
Suspendisse aliquet, sem ut cursus luctus, ipsum	3057.3
ac libero nec ligula consectetuer	2972.31
non,	2926.17
Sed diam lorem, auctor quis, tristique	2851.42

id ante dictum 2786.32

in, dolor. Fusce 2777

FLOORLOCATION TOTALSALES

---------- ----------

a felis ullamcorper viverra. 2715.11

Curae Phasellus ornare. 2668.4

facilisis lorem tristique aliquet. Phasellus ferme 2522.8

tempor diam dictum sapien. Aenean massa. Integer v 2430.12

Sed id risus quis 2380.2

27 rows selected.

What are the total sales in a given month?

The class used in this question is the Purchase table, using the columns PurchaseDate and TotalCost. From the PurchaseDate, we will find the month and then sum all the total costs by month. This information will be useful to the higher ups of the company because it will show when traffic and purchasing at the company is most high, which can lead to improved stocking and more dynamic staffing to match the influxes and decreases. It also may lead into questions of if a certain month has high purchases, what products are being purchased.

Code:

```
DECLARE
  total_sales DECIMAL(11,2);
  error_message VARCHAR2(200);
BEGIN

  SELECT ROUND(SUM(TotalCost)) INTO total_sales
```

```
 FROM Purchase
 WHERE PurchaseDate >= '9/01/12' AND PurchaseDate <= '9/31/12';
 DBMS_OUTPUT.PUT_LINE('The total sales for September 2012 is: ' || total_sales);

 EXCEPTION
  WHEN NO_DATA_FOUND THEN
    error_message := 'No records found!';
    DBMS_OUTPUT.PUT_LINE(error_message);
  WHEN OTHERS THEN
    error_message := SQLERRM;
    DBMS_OUTPUT.PUT_LINE('The error that occurred: ' || error_message);
END;
```

Output:

```
The total sales for September 2012 is: 6319
```

What is the average amount of rewards a Target customer receives as a bronze tier member?

This question will use the Purchase table and use the PointsEarned column, as well as the Rewards table using the RewardsLevel column. From this, we can calculate the average of all earned points per purchase. This is beneficial because the overall average can be a good reflection of purchasing trends of bronze tier members. This would be of interest for the marketing team in seeing if bronze members are earning rewards enough to keep them incentivized by the rewards program, since bronze is the entry level rewards program.

Code:

```
SELECT ROUND(AVG(Rewards.Points)) AS Avg_Rewards
FROM Customer1
JOIN Rewards ON Customer1.RewardsID = Rewards.RewardsID
WHERE Rewards.RewardsLevel = 'Bronze';
```

Output:

```
 AVG_REWARDS
-----------
         498
```

What month has the most rewards used by rewards level, and how many?

The classes used to answer this question are Purchase and Rewards. In purchase, we use the PurchaseDate, UsedPoints and in Rewards, we use RewardsLevel. These questions would provide interesting insight to what month for each tier level are the most points being used. This kind of information would be useful for people in charge of the rewards program to see point usage by date and tier to see if reward incentives can be placed on different days of the week to increase point usage, therefore store purchase.

Code:

```
SELECT
    TO_CHAR(TO_DATE(SUBSTR(PurchaseDate, 1, 10), 'MM/DD/YY'), 'Month') AS MonthName,
    RewardsLevel,
    COUNT(RewardsUsed) AS RewardsUsed
FROM Purchase
JOIN Rewards ON Purchase.RewardsID = Rewards.RewardsID
GROUP BY TO_CHAR(TO_DATE(SUBSTR(PurchaseDate, 1, 10), 'MM/DD/YY'), 'Month'),
RewardsLevel
ORDER BY RewardsUsed DESC;
```

Output:

```
 MONTHNAME REWARDSLEVEL
REWARDSUSED
--------- ------------------------- -----------
October   Gold
13
November  Silver
9
```

January Silver
9

January Bronze
8

November Gold
8

August Bronze
8

December Gold
8

May Bronze
8

December Bronze
8

December Silver
7

November Bronze
7

MONTHNAME REWARDSLEVEL
REWARDSUSED

--------- --------------------------- -----------

May Silver
6

May Gold
6

March Silver
6

September Silver
6

July Gold
6

February Silver
6

September Bronze
5

October Silver
5

January Gold
5

August Silver
5

March Bronze
5

MONTHNAME REWARDSLEVEL
REWARDSUSED

--------- ---------- -----------

September Gold
5

June Gold
4

March Gold
4

February Bronze
4

July Silver
4

February Gold
3

August Gold
3

April Bronze
3

June Bronze
3

April Silver
3

July Bronze
2

MONTHNAME REWARDSLEVEL
REWARDSUSED

--------- ------------------ -----------

April Gold
2

October Bronze
2

June Silver
2

36 rows selected.

What floor locations have the most purchase?

For this question, we will use the Purchase table with the columns TotalCost and Zip. We will add up all total costs and compare it with all other zip codes. This will be useful to find the more successful stores or the most successful areas. This will show purchasing trends within geographical areas like in cities or suburbs. We can use this info to determine where other stores should be built.

Code:

```
SELECT p.floorlocation, SUM(pp.quantity) AS total_purchases
FROM Product p
INNER JOIN Purchase_Products pp ON p.productID = pp.productID
GROUP BY p.floorlocation
ORDER BY total_purchases DESC;
```

Output:

FLOORLOCATION TOTAL_PURCHASES

---------- ---------- ---------- ----------

Donec est 76
facilisis lorem tristique aliquet. Phasellus ferme 70
a felis ullamcorper viverra. 61
et, commodo 53
sagittis semper. 52
dolor dapibus 52
at pede. Cras vulputate velit eu sem. 48
enim. Nunc ut erat. Sed nunc est, mollis 47
nibh enim, gravida sit 47
Curabitur sed tortor. Integer aliquam 46
Nullam nisl. Maecenas malesuada 45

FLOORLOCATION	TOTAL_PURCHASES
et, lacinia vitae, sodales at, velit. Pellentesque	45
a	44
tristique neque venenatis lacus. Etiam	43
Aliquam vulputate ullamcorper magna. Sed eu eros.	43
scelerisque neque. Nullam	43
lectus pede, ultrices a, auctor non, feugiat	41
et ultrices posuere cubilia Curae Phasellus ornare	40
Curae Phasellus ornare.	40
eu nibh vulputate mauris sagittis	39
parturient montes, nascetur ridiculus mus. Proin v	39
non,	39

FLOORLOCATION	TOTAL_PURCHASES
auctor ullamcorper, nisl arcu	38

sit amet, dapibus id, blandit at, nisi. Cum sociis 38
semper, dui lectus rutrum urna, nec luctus felis p 37
mattis. Cras eget nisi dictum augue malesuada male 37
adipiscing. Mauris 36
In at pede. 36
ante. Maecenas mi felis, adipiscing fringilla, 36
a, arcu. Sed et libero. Proin 35
sollicitudin adipiscing ligula. Aenean gravida nun 35
lorem tristique aliquet. Phasellus 34
felis. 34

FLOORLOCATION	TOTAL_PURCHASES
velit. Pellentesque ultricies dignissim lacus.	34
Phasellus nulla. Integer vulputate, risus	33
ornare, lectus	33
ipsum primis in faucibus orci luctus et ultrices p	33
ac libero nec ligula consectetuer	33
Nulla tincidunt,	33
elit sed consequat auctor, nunc nulla vulputate du	33
tincidunt tempus risus. Donec egestas. Duis ac arc	32
diam. Pellentesque habitant morbi	32
vel arcu. Curabitur ut odio vel est tempor	32
dolor sit amet, consectetuer adipiscing elit. Cura	31

FLOORLOCATION	TOTAL_PURCHASES
in faucibus orci luctus et ultrices posuere cubili	31
placerat velit. Quisque	30

turpis vitae purus gravida sagittis. Duis 30

Suspendisse aliquet, sem ut cursus luctus, ipsum 30

Sed id risus quis 30

netus et malesuada fames ac turpis egestas. Fusce 30

mollis. Duis sit amet diam 30

sem eget massa. Suspendisse eleifend. Cras sed leo 29

at pretium aliquet, metus urna convallis erat, ege 29

non arcu. Vivamus sit amet risus. 29

id ante dictum 29

Who are the top earning employees within their respective job title?

For this question, we will use the Employee table with the columns Pay, EmployeeID, and JobTitle. We will select those columns and find the highest meaning employee, and sort it by jobtitle. This will be interesting and useful information as we would be able to see the differences in pay between employees. This can show it someone is over or underpaid.

Code:

```
SELECT EmpFirstName, EmpLastName, JobTitle, Pay
FROM (
   SELECT EmpFirstName, EmpLastName, JobTitle, Pay, RANK() OVER (PARTITION BY JobTitle ORDER BY Pay DESC) AS pay_rank
   FROM Employees
) ranked_employees
WHERE pay_rank = 1;
```

Output:

EMPFIRSTNAME	EMPLASTNAME	JOBTITLE	PAY

Flynn	Holt	Cashier	10
Derek	Heath	Cashier	10
Felix	Riley	Cashier	10
Donovan	Barrera	Cashier	10
Burke	Sparks	Cashier	10
Xenos	Hall	Cashier	10
Ciaran	Mclean	Cashier	10
Isaiah	Baldwin	Cashier	10
Lee	Rodgers	Cashier	10
Brett	Levine	Cashier	10
Zeus	Woodward	General Manager	22

EMPFIRSTNAME	EMPLASTNAME	JOBTITLE	PAY
Alfonso	Gardner	Janitor	9.25
Tad	Benson	Janitor	9.25
Tanek	Rodriguez	Janitor	9.25
Benjamin	Downs	Janitor	9.25
Edan	Fields	Janitor	9.25
Oliver	Gutierrez	Janitor	9.25
Elton	Orr	Janitor	9.25
Todd	Rosario	Janitor	9.25
Jacob	Buchanan	Janitor	9.25
Mufutau	Middleton	Janitor	9.25
Isaiah	Frost	Manager	16

EMPFIRSTNAME	EMPLASTNAME	JOBTITLE	PAY

Jerome	Church	Manager	16
Davis	Mcdowell	Manager	16
Judah	Powell	Manager	16
Cole	Pacheco	Manager	16
Cairo	Scott	Manager	16
Nicholas	Moran	Stocker	11
Kasper	Clark	Stocker	11
Dylan	Parsons	Stocker	11
Dean	Frank	Stocker	11
Tanek	Contreras	Stocker	11
Damian	Stephens	Stocker	11

EMPFIRSTNAME	EMPLASTNAME	JOBTITLE	PAY
Paul	Barrett	Stocker	11
Todd	Jarvis	Stocker	11
Davis	Maxwell	Stocker	11
Cullen	Mcdowell	Stocker	11
Ethan	Dyer	Stocker	11
Hop	Lloyd	Stocker	11
Chester	Cunningham	Stocker	11
Griffin	Higgins	Stocker	11
Hamish	Baker	Stocker	11
Samson	Mcleod	Stocker	11
Jakeem	Orr	Team Lead	13

EMPFIRSTNAME	EMPLASTNAME	JOBTITLE	PAY
Harper	Leblanc	Team Lead	13

Gil	Brown	Team Lead	13
Mason	Clayton	Team Lead	13
Zahir	Sharpe	Team Lead	13
Isaac	Shields	Team Lead	13
Acton	Mitchell	Team Lead	13

50 rows selected.

What percentage of customers have gold tier rewards?

For this question, we will use the customer table and the rewards table which join on the RewardsID column. Additionally, we will utilize the count function to calculate the percentage of customers that have 'gold' as their rewards level in the RewardsLevel column. This information will be useful to those with higher positions in the company to track the rewards in search of areas to improve customer loyalty.

Code:

```
SELECT
   (COUNT(CASE WHEN Rewards.RewardsLevel = 'Gold' THEN 1 END) * 100.0 / COUNT(*)) AS GoldRewardsPercent
FROM Customer1
LEFT JOIN Rewards ON Customer1.RewardsID = Rewards.RewardsID;
```

Output:

```
 GOLDREWARDSPERCENT
------------------
       32.8976035
```

How many rewards were started during the Winter Holiday's season compared to the summer season?

In order to obtain information about the rewardstartdate, we will use the rewards table for this question. To ensure that our search is properly focused on the relevant information, we will use the case when function. The rewardstartdate varies in different holiday seasons

such as Thanksgiving and Christmas, as compared to the summer season. This information will allow for more effective marketing strategies to be implemented in order to maximize sales during the appropriate seasons. It is crucial to understand the differences in sales patterns and trends to determine how much budget to allocate for marketing in each specific season.

Code:

```
SELECT
    SUM(CASE WHEN RewardStartDate LIKE '11%' OR RewardStartDate LIKE '12%' THEN 1 ELSE 0 END) AS ChristmasSignups,
    SUM(CASE WHEN RewardStartDate LIKE '6%' OR RewardStartDate LIKE '7%' OR RewardStartDate LIKE '8%' THEN 1 ELSE 0 END) AS SummerSignups
FROM Rewards;
```

Output:

```
 CHRISTMASSIGNUPS SUMMERSIGNUPS
---------------- -------------
              77           136
```

Team

I. Jenna Lusk

A majority of my work was creating the data. This was a very large task because I had to randomly generate many columns, but then ensure FK and PK were used appropriately in order to perform some of the questions we wished to answer. I also organized the final report document to ensure every part was covered. I created the file structure of the github repo. I also provided a solution to one of the questions.

II. Rose Elder

The majority of my was creating answers and solutions. Specifically I created questions 6, 7, and 8. I then created solutions for questions 3, 7, and 8. I then created the readme.md that describes the theme and goal of our project, as well as instructions for using our code. I also assisted in creating the original outlines of our tables and conversed with my fellow teammates in creating our goal for our project.

III. Khushi Saini

Initially, my primary focus was on creating a GitHub repository for the project and developing answers and solutions to the questions we had formulated. Among the questions that were assigned to me, I took responsibility for generating questions 4, 5, and 10 and provided the relevant code and solutions for each. Moreover, I collaborated with my team members to develop the Background section of our final report, and we worked together to refine our vision and unify our approach to the project.

With that being said, my contributions to this project spanned several areas, including creating the GitHub repository, generating and solving assigned questions, and collaborating with my team to develop the Background section of our final report. Through active participation in the group, I played a vital role in achieving our collective goal and contributing to the overall success of the project along with all our teammates.

IV. Jennifer Merry

The majority of my work was brainstorming questions and coding the solutions. I developed solutions for questions 1, 6, and 9 and I created the descriptions for questions 8 and 9. I also wrote the ‘background’ section of the final report.

Additionally, I aided my teammates in brainstorming our scenario and the tables and columns that we would use.