1. **How many Customers do we have in the data?**

Query 1:

SELECT

COUNT(DISTINCT customer\_id)

FROM

customers

**result:795**

1. **What was the city with the most profit for the company in 2015 and how much was it?**

Query 2:

SELECT

shipping\_city,

SUM(order\_profits),

order\_date

FROM

orders

JOIN order\_details ON orders.order\_id = order\_details.order\_id

WHERE

order\_date LIKE '%2015%'

GROUP BY

1

ORDER BY

2 DESC

LIMIT

5

**result:**

|  |  |  |
| --- | --- | --- |
| **shipping\_city** | **SUM(order\_profits)** | **order\_date** |
| New York City | 14753 | 9/6/2015 |
| Seattle | 5071 | 12/7/2015 |
| Minneapolis | 4695 | 9/23/2015 |
| San Francisco | 4290 | 7/8/2015 |
| Los Angeles | 4092 | 4/20/2015 |

New York City 14753

1. **How many different cities do we have in the data?**

Query 3:

SELECT

COUNT(DISTINCT shipping\_city)

FROM

Orders

**result: 531**

1. **Show the total spent by customers from low to high.**

Query 4:

SELECT

customers.customer\_id,

SUM(order\_sales) total\_spent\_customers

FROM

customers

JOIN orders ON customers.customer\_id = orders.customer\_id

JOIN order\_details ON orders.order\_id = order\_details.order\_id

GROUP BY

customers.customer\_id

ORDER BY

total\_spent\_customers ASC

**result:**

|  |  |
| --- | --- |
| **customer\_id** | **total\_spent\_customers** |
| 456 | 5 |
| 738 | 5 |
| 546 | 16 |
| 124 | 17 |
| 657 | 22 |

1. **What is the most profitable City in the State of Tennessee?**

Query 5:

SELECT

shipping\_city,

AVG (order\_profits)

FROM

order\_details

JOIN orders USING (order\_id)

WHERE

shipping\_state = 'Tennessee'

GROUP BY

shipping\_city

ORDER BY

AVG (order\_profits) DESC

**result:**

|  |  |
| --- | --- |
| **shipping\_city** | **avg (order\_profits)** |
| Lebanon | 27.6666666666667 |
| Bartlett | 10.0 |
| Smyrna | 2.85714285714286 |
| Chattanooga | 2.8 |
| Murfreesboro | 1.36363636363636 |
| Johnson City | 1.0 |
| Jackson | -4.4 |
| Franklin | -8.45454545454546 |
| Hendersonville | -9.2 |
| Bristol | -25.4 |
| Nashville | -28.8965517241379 |
| Columbia | -36.9444444444444 |
| Knoxville | -48.625 |
| Memphis | -49.3 |
| Clarksville | -150.857142857143 |

1. **What’s the average annual profit for that city across all years in that city?**

Query 6:

SELECT

AVG(order\_profits) AS average,

shipping\_city

FROM

order\_details

JOIN orders USING (order\_id)

WHERE

shipping\_city = 'Lebanon'

**result:**

|  |  |
| --- | --- |
| **average** | **shipping\_city** |
| 27.6666666666667 | Lebanon |

1. **What is the distribution of customer types in the data?**

Query 7:

SELECT

customer\_segment,

COUNT(customer\_id) AS total\_customers

FROM

customers

GROUP BY

customer\_segment

**result:**

|  |  |
| --- | --- |
| Consumer | 410 |
| Corporate | 237 |
| Home Office | 148 |

1. **What’s the most profitable product category on average in Iowa across all years?**

Query 8:

SELECT

product\_category,

AVG(order\_profits) AS most\_profit

FROM

product

JOIN order\_details ON product.product\_id = order\_details.product\_id

JOIN orders ON order\_details.order\_id = orders.order\_id

WHERE

shipping\_state = 'Iowa'

GROUP BY

product\_category

**result:**

|  |  |
| --- | --- |
| Furniture | 130.25 |
| Office Supplies | 15.72 |
| Technology | 79.75 |

1. **What is the most popular product in that category across all states in 2016?**

Query 9:

SELECT

product\_name,

product\_category,

SUM(quantity)

FROM

product

JOIN order\_details ON product.product\_id = order\_details.product\_id

JOIN orders ON order\_details.order\_id = orders.order\_id

WHERE

order\_date LIKE '%2016'

AND product\_category = 'Furniture'

GROUP BY

product\_name

ORDER BY

SUM(quantity) DESC

LIMIT

1

**result:**

|  |  |  |
| --- | --- | --- |
| **product\_name** | **product\_category** | **Sum(quantity)** |
| Global Push Button Manager's Chair, Indigo | Furniture | 22 |

1. **Which customer got the most discount in the data? (In total amount)**

Query 10:

SELECT

c.customer\_id AS customer,

customer\_name,

SUM((order\_sales / (1 - order\_discount)) - order\_sales) AS total\_discount

FROM

orders AS o

JOIN order\_details AS od ON o.order\_id = od.order\_id

JOIN customers AS c ON o.customer\_id = c.customer\_id

GROUP BY

customer,

customer\_name

ORDER BY

total\_discount DESC

LIMIT

5;

**result:**

|  |  |  |
| --- | --- | --- |
| **customer** | **customer\_name** | **total\_discount** |
| 687 | Sean Miller | 23929.0833333333 |
| 166 | Cindy Stewart | 11594.3333333333 |
| 478 | Luke Foster | 9052.16666666667 |
| 308 | Grant Thornton | 8746.83333333334 |
| 330 | Henry Goldwyn | 7692.58333333334 |

1. **How widely did monthly profits vary in 2018?**

Query 11:

WITH y2018 AS(

SELECT

substr(o.order\_date, length(o.order\_date) -3) AS YEAR,

CAST(

substr(o.order\_date, 0, instr(o.order\_date, '/')) AS INT

) AS MONTH,

SUM(d.order\_profits) AS profit

FROM

orders o

LEFT JOIN order\_details d ON o.order\_id = d.order\_id

WHERE

YEAR = "2018"

GROUP BY

1,

2

ORDER BY

2

)

SELECT

\*,

lag(profit) OVER(

ORDER BY

MONTH

) AS preceding\_profit,

profit - lag(profit) OVER(

ORDER BY

MONTH

) AS delta\_profit

FROM

y2018

**result:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **YEAR** | **MONTH** | **profit** | **preceding\_profit** | **delta\_profit** |
| 2018 | 1 | 7137 |  |  |
| 2018 | 2 | 1612 | 7137 | -5525 |
| 2018 | 3 | 14758 | 1612 | 13146 |
| 2018 | 4 | 934 | 14758 | -13824 |
| 2018 | 5 | 6342 | 934 | 5408 |
| 2018 | 6 | 8226 | 6342 | 1884 |
| 2018 | 7 | 6951 | 8226 | -1275 |
| 2018 | 8 | 9034 | 6951 | 2083 |
| 2018 | 9 | 10987 | 9034 | 1953 |
| 2018 | 10 | 9272 | 10987 | -1715 |
| 2018 | 11 | 9217 | 9272 | -55 |
| 2018 | 12 | 8473 | 9217 | -744 |

1. **Which order was the highest in 2015?**

Query 12:

SELECT

MAX(order\_sales) AS highest\_order,

order\_id

FROM

order\_details

JOIN orders USING(order\_id)

WHERE

order\_date LIKE '%2015'

**result:**

|  |  |
| --- | --- |
| **highest\_order** | **order\_id** |
| 22638 | CA-2015-145317 |

1. **What was the rank of each city in the East region in 2015?**

Query 13:

SELECT

DISTINCT o.shipping\_city,

od.quantity,

RANK() OVER (

ORDER BY

SUM(od.quantity) desc

) AS ranking

FROM

order\_details AS od

JOIN orders AS o USING(order\_id)

WHERE

o.order\_date LIKE '%2015'

AND o.shipping\_region LIKE 'East'

GROUP BY

o.shipping\_city

**result:**

|  |  |  |
| --- | --- | --- |
| **shipping\_city** | **quantity** | **ranking** |
| New York City | 62 | 1 |
| Philadelphia | 3 | 2 |
| Columbus | 3 | 3 |
| Newark | 4 | 4 |
| Fairfield | 7 | 5 |
| Long Beach | 2 | 6 |
| Lakewood | 7 | 7 |
| Lancaster | 6 | 8 |
| Lawrence | 3 | 9 |
| Dover | 2 | 10 |

1. **Join all DB tables into one dataset that includes all unique columns and download it as a CSV file. In the second part of the project, you're gonna work with this one table.**

Query 14:

SELECT

\*

FROM

customers

JOIN orders ON orders.customer\_id = customers.customer\_id

JOIN order\_details ON order\_details.order\_id = orders.order\_id

JOIN product ON product.product\_id = order\_details.product\_id