**CAPSTONE PROJECT**

**Dataset explanation:**

The online sector has been slowly eating up market share in the past two decades. E-commerce platforms like Unicorn allow people to buy products online: from books, toys, clothes, and shoes to food, furniture, and other household items. The following dataset includes Unicorn sales data from the years 2015-2018.

**Dataset link:**

https://bit.io/emiliano/Capstone\_Project

**Questions:**

1. How many customers do we have in the data?

Code:

SELECT COUNT(customer\_id)

FROM customers;

A: *795*

1. What was the city with the most profit for the company in 2015?

Code:

SELECT SUM(order\_profits), shipping\_city,

DATE\_PART('year',order\_date) AS year

FROM orders

JOIN order\_details

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

WHERE DATE\_PART('year',order\_date) = 2015

GROUP BY 2, 3

ORDER BY 1 DESC

LIMIT 1;

A: New York City

1. In 2015, what was the most profitable city's profit?

Code: Same with Q2.

A: 14753

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

Code:

SELECT COUNT(DISTINCT(shipping\_city))

FROM orders;

A: 531

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

Code:

SELECT customer\_name, o.customer\_id, SUM(order\_sales) AS total\_spent

FROM customers c

JOIN orders o

ON c.customer\_id = o.customer\_id

JOIN order\_details od

ON o.order\_id = od.order\_id

GROUP BY 1, 2

ORDER BY 3;

A: 456

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

Code:

SELECT SUM(order\_profits) total\_profit, shipping\_city, shipping\_state

FROM orders o

JOIN order\_details od

ON o.order\_id = od.order\_id

WHERE shipping\_state = 'Tennessee'

GROUP BY 2,3

ORDER BY 1 DESC

LIMIT 1;

A: Lebanon

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

Code:

SELECT AVG(order\_profits)::numeric(10,2) total\_profit, shipping\_city

FROM order\_details od

JOIN orders o

ON od.order\_id = o.order\_id

WHERE shipping\_city = 'Lebanon'

GROUP BY 2;

A: 27.67

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

Code:

SELECT COUNT(customer\_id), customer\_segment

FROM customers

GROUP BY 2;

A:

|  |  |
| --- | --- |
| **count** | **customer\_segment** |
| **410** | Consumer |
| **237** | Corporate |
| **148** | Home Office |

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

Code:

SELECT AVG(order\_profits)::numeric(10,2) avg\_profit, product\_category, shipping\_state

FROM orders o

JOIN order\_details od

ON o.order\_id = od.order\_id

JOIN product p

ON p.product\_id = od.product\_id

WHERE shipping\_state = 'Iowa'

GROUP BY 2, 3

ORDER BY 1 DESC

LIMIT 1;

A: Furniture

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

Code:

SELECT COUNT(o.order\_id) total\_orders, product\_name,

product\_category

FROM orders o

JOIN order\_details od

ON o.order\_id = od.order\_id

JOIN product p

ON p.product\_id = od.product\_id

WHERE product\_category = 'Furniture'

AND DATE\_PART('year', order\_date) = 2016

GROUP BY 2, 3

ORDER BY 1 DESC;

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

Code:

SELECT customer\_name, c.customer\_id, SUM(order\_discount)::numeric(10,2) total\_discount

FROM customers c

JOIN orders o

ON o.customer\_id = c.customer\_id

JOIN order\_details od

ON od.order\_id = o.order\_id

GROUP BY 1, 2

ORDER BY 3 DESC

LIMIT 1;

A: Sean Miller, 687

1. How widely did monthly profits vary in 2018?

Code:

SELECT SUM(order\_profits)::numeric(10,2) total\_profits,

DATE\_PART('month', order\_date) AS month

FROM order\_details od

JOIN orders o

ON o.order\_id = od.order\_id

WHERE DATE\_PART('year', order\_date) = 2018

GROUP BY 2

ORDER BY 2;

* Add difference column with lag

A: There is huge decline in February and April.

|  |  |
| --- | --- |
| **total\_profits** | **month** |
| **7137.00** | 1 |
| **1612.00** | 2 |
| **14758.00** | 3 |
| **934.00** | 4 |
| **6342.00** | 5 |
| **8226.00** | 6 |
| **6951.00** | 7 |
| **9034.00** | 8 |
| **10987.00** | 9 |
| **9272.00** | 10 |
| **9217.00** | 11 |
| **8473.00** | 12 |

1. Which order was the highest in 2015?

Code:

SELECT o.order\_id, quantity, order\_profits

FROM order\_details od

JOIN orders o

ON o.order\_id = od.order\_id

WHERE DATE\_PART('year', order\_date) = 2015

--ORDER BY 3 DESC;

ORDER BY 2 DESC;

A: Highest quantity: CA-2015-100136 with 324

Highest profit: CA-2015-116904 with 4630

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

Code:

SELECT SUM(quantity) order\_count, shipping\_city

FROM order\_details od

JOIN orders o

ON o.order\_id = od.order\_id

WHERE DATE\_PART('year', order\_date) = 2015

AND shipping\_region = 'East'

GROUP BY 2

ORDER BY 1 DESC;

* ADD rank column

A: New York first in order count.

1. Display customer names for customers who are in the segment ‘Consumer’ or ‘Corporate.’ How many customers are there in total?

Code:

SELECT customer\_name, customer\_segment

FROM customers

WHERE customer\_segment = 'Consumer'

OR customer\_segment = 'Corporate'

A: 795

1. Calculate the difference between the largest and smallest order quantities for product id ‘100.’

Code:

SELECT MAX(quantity) - MIN(quantity) difference

FROM order\_details

WHERE product\_id = '100'

A: 4

1. Calculate the percent of products that are within the category ‘Furniture.’

Code:

SELECT ((

SELECT COUNT(product\_id)

FROM product

WHERE product\_category = 'Furniture'

) \* 100.0 / COUNT(product\_id))::numeric(10,2) AS percentage

FROM product

A: %20.54

1. Display the number of duplicate products based on their product manufacturer. Example: A product with an identical product manufacturer can be considered a duplicate.

Code:

SELECT COUNT(\*), product\_manufacturer

FROM product

GROUP BY 2;

1. Show the product\_subcategory and the total number of products in the subcategory. Show the order from most to least products and then by product\_subcategory name ascending.

Code:

SELECT product\_subcategory, COUNT(product\_subcategory) total\_product\_count

FROM product

group by 1

ORDER BY 2 DESC, 1;

A: Paper 277

1. Show the product\_id(s), the sum of quantities, where the total sum of its product quantities is greater than or equal to 100.

Code:

SELECT product\_id, quantity

FROM order\_details

WHERE quantity >= 100

**Bonus question:**

1. Join all database tables into one dataset that includes all unique columns and download it as a .csv file.

Code:

SELECT \*

FROM customers c

JOIN orders o

ON c.customer\_id = o.customer\_id

JOIN order\_details od

ON o.order\_id = od.order\_id

JOIN product p

ON od.product\_id = p.product\_id;

A:

**Codes Written**