**DA104.5 Mini-Project**

**Data Exploration with SQL**

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

| 1  2  3  4 | **SELECT**  **COUNT**(**DISTINCT** customer\_id) **AS** total\_customers  **FROM**  customers; |
| --- | --- |

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

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

| 1  2  3  4  5  6  7  8  9  10 | **SELECT**  ord.shipping\_city **as** city,  **SUM**(det.order\_profits) **AS** profit  **FROM** orders ord  **JOIN** order\_details det **ON** ord.order\_id = det.order\_id  **WHERE**  **EXTRACT**(**YEAR** **FROM** ord.order\_date) = **2015**  **GROUP** **BY** **1**  **ORDER** **BY** **2** **DESC**  **LIMIT** **1**; |
| --- | --- |

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

| 1  2  3 | **SELECT**  **COUNT**(**DISTINCT** shipping\_city) **AS** total\_unique\_cities  **FROM** orders; |
| --- | --- |

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

| 1  2  3  4  5  6  7  8  9 | **SELECT**  ord.customer\_id,  cu.customer\_name,  **SUM**(det.order\_sales) **AS** total\_spent  **FROM** orders ord  **JOIN** order\_details det **ON** ord.order\_id = det.order\_id  **JOIN** customers cu **ON** ord.customer\_id = cu.customer\_id  **GROUP** **BY** **1**, **2**  **ORDER** **BY** **3** **ASC**; |
| --- | --- |

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

| 1  2  3  4  5  6  7  8  9  10 | **SELECT**  ord.shipping\_city,  **SUM**(det.order\_profits) **AS** total\_profit\_city  **FROM** orders ord  **JOIN** order\_details det **ON** ord.order\_id = det.order\_id  **WHERE**  ord.shipping\_state = 'Tennessee'  **GROUP** **BY** **1**  **ORDER** **BY** **2** **DESC**  **LIMIT** **1**; |
| --- | --- |

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

| 1  2  3  4  5  6  7  8  9  10  11  12 | **WITH** temp\_tab **AS**(  **SELECT**  **EXTRACT**(**YEAR** **FROM** ord.order\_date) **AS** years,  **AVG**(det.order\_profits) **AS** average\_profit  **FROM** orders ord  **JOIN** order\_details det **ON** ord.order\_id = det.order\_id  **WHERE**  ord.shipping\_city = 'Lebanon'  **GROUP** **BY** **1**  **ORDER** **BY** **2** **DESC**)  **SELECT** **AVG**(average\_profit) **AS** average\_annual\_profit  **FROM** temp\_tab; |
| --- | --- |

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

| 1  2  3  4 | **SELECT** **DISTINCT** customer\_segment **AS** customer\_types,  **COUNT**(customer\_segment) **AS** cnt  **FROM** customers  **GROUP** **BY** **1**; |
| --- | --- |

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

| 1  2  3  4  5  6  7  8  9  10 | **SELECT**  pr.product\_category,  **AVG**(det.order\_profits) **AS** average\_profit  **FROM** product pr  **JOIN** order\_details det **ON** pr.product\_id = det.product\_id  **JOIN** orders ord **ON** ord.order\_id = det.order\_id  **WHERE** ord.shipping\_state = 'Iowa'  **GROUP** **BY** **1**  **ORDER** **BY** **2** **DESC**  **LIMIT** **1**; |
| --- | --- |

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

| 1  2  3  4  5  6  7  8  9  10  11  12 | **SELECT**  pr.product\_id,  pr.product\_name,  **SUM**(det.quantity) **AS** order\_qty  **FROM** product pr  **JOIN** order\_details det **ON** pr.product\_id = det.product\_id  **JOIN** orders ord **ON** ord.order\_id = det.order\_id  **WHERE** product\_category = 'Furniture'  **AND** **EXTRACT**(**YEAR** **FROM** ord.order\_date) = **2016**  **GROUP** **BY** **1**, **2**  **ORDER** **BY** **3** **DESC**  **LIMIT** **1**; |
| --- | --- |

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

| 1  2  3  4  5  6  7  8  9  10  11  12  13  14 | **WITH** temp\_tab **AS**(  **SELECT**  ord.customer\_id **AS** customer\_id,  **CASE** **WHEN** det.order\_discount != **0** **THEN** (det.order\_sales \*  det.order\_discount) / (**1** - det.order\_discount)  **ELSE** order\_discount  **END** **AS** total\_discount  **FROM** orders ord  **JOIN** order\_details det **ON** ord.order\_id = det.order\_id)  **SELECT** customer\_id,  **SUM**(total\_discount) **AS** total\_discount  **FROM** temp\_tab  **GROUP** **BY** **1**  **ORDER** **BY** **2** **DESC**; |
| --- | --- |

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

| 1  2  3  4  5  6  7  8  9  10  11    12    13  14 | **WITH** temp\_tab **AS**(  **SELECT**  **EXTRACT**(**MONTH** **FROM** ord.order\_date) **AS** order\_month,  **SUM**(det.order\_profits) **AS** profit\_per\_month  **FROM** orders ord  **JOIN** order\_details det **ON** ord.order\_id = det.order\_id  **WHERE** **EXTRACT**(**YEAR** **FROM** ord.order\_date) = **2018**  **GROUP** **BY** **1**)  **SELECT** order\_month, profit\_per\_month,  **CASE** **WHEN**  LAG(profit\_per\_month, **1**, **0**) OVER(**ORDER** **BY** order\_month) = **0** **THEN**  profit\_per\_month  **ELSE** (profit\_per\_month - LAG(profit\_per\_month, **1**, **0**) OVER(**ORDER** **BY**  order\_month))  **END** **AS** profit\_diff\_per  **FROM** temp\_tab; |
| --- | --- |

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

| 1  2  3  4  5  6  7  8  9  10 | **SELECT**  ord.order\_id,  **SUBSTRING**(ord.order\_id, **9**, **6**) **AS** order\_id\_l,  **SUM**(det.quantity) **AS** total\_qty,  **SUM**(DET.order\_sales) **AS** total\_sales  **FROM** orders ord  **JOIN** order\_details det **ON** ord.order\_id = det.order\_id  **WHERE** **EXTRACT**(**YEAR** **FROM** ord.order\_date) = **2015**  **GROUP** **BY** **1**  **ORDER** **BY** **4** **DESC**; |
| --- | --- |

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

| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16 | **WITH**  temp\_tab **AS** (  **SELECT**  ord.shipping\_city,  **SUM**(det.quantity) **AS** qty  **FROM** orders ord  **JOIN** order\_details det **ON** ord.order\_id = det.order\_id  **WHERE**  **EXTRACT**(**YEAR** **FROM** ord.order\_date) = **2015**  **AND** ord.shipping\_region = 'East'  **GROUP** **BY** **1**)  **SELECT**  shipping\_city,  qty,  RANK() OVER (**ORDER** **BY** qty **DESC**) **AS** rank\_city  **FROM** temp\_tab; |
| --- | --- |

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

| 1  2  3  4 | **SELECT**  **COUNT**(**DISTINCT** customer\_name) **AS** total\_customer  **FROM** customers  **WHERE** customer\_segment **IN** ('Consumer', 'Corporate'); |
| --- | --- |

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

| 1  2  3  4  5 | **SELECT**  **MAX**(quantity), **MIN**(quantity),  **MAX**(quantity) - **MIN**(quantity) **AS** difference  **FROM** order\_details  **WHERE** product\_id = '100'; |
| --- | --- |

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

| 1  2  3  4 | **SELECT**  (**COUNT**(**CASE** **WHEN** product\_category = 'Furniture' **THEN** **1** **END**)::numeric / **COUNT**(\*)) \* **100** **AS** percent\_furniture  **FROM** product; |
| --- | --- |

**18. Display the number of product manufacturers with more than 1 product in the product table.**

**Example: A product with an identical product manufacturer can be considered a duplicate.**

| 1  2  3  4  5  6 | **SELECT**  product\_manufacturer,  **COUNT**(\*) **AS** num\_of\_duplicates  **FROM** product  **GROUP** **BY** **1**  **HAVING** **COUNT**(\*) > **1**; |
| --- | --- |

**19. 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.**

| 1  2  3  4  5  6 | **SELECT**  product\_subcategory,  **COUNT**(product\_id) **AS** products  **FROM** product  **GROUP** **BY** **1**  **ORDER** **BY** **2** **DESC**, **1** **ASC**; |
| --- | --- |

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

| 1  2  3  4  5  6  7 | **SELECT**  product\_id,  **SUM**(quantity) **AS** total\_qty  **FROM** order\_details  **WHERE** quantity >= **100**  **GROUP** **BY** **1**  **ORDER** **BY** **2** **ASC**; |
| --- | --- |

**\*\*Bonus question:\*\* -- Join all database tables into one dataset that includes all unique columns and download it as a .csv file.**

| 1  2  3  4  5 | **SELECT** \*  **FROM** customers cu  **JOIN** orders ord **ON** cu.customer\_id = ord.customer\_id  **JOIN** order\_details det **ON** ord.order\_id = det.order\_id  **JOIN** product pr **ON** det.product\_id = pr.product\_id; |
| --- | --- |

For unique columns

| 1  2  3  4  5 | **SELECT** \*  **FROM** customers  **NATURAL** **JOIN** orders  **NATURAL** **JOIN** order\_details  **NATURAL** **JOIN** product; |
| --- | --- |