DSCI 504: Intro to SQL

Assignment 07: Data Transformations, Subqueries, and Window Functions

Jeffery Boczkaja

**A.--------------------------------------------------------------------------------**

A screenshot of a computer

Description automatically generated with medium confidence

i.A screenshot of a computer screen

Description automatically generated with low confidence

ii. A screenshot of a computer screen

Description automatically generated with low confidence

There were 138 orders in the year 2010.

iii.A screenshot of a computer

Description automatically generated with medium confidence

iv.A screenshot of a computer

Description automatically generated with medium confidence

v. A screenshot of a computer

Description automatically generated with medium confidence

The year with the most customers was 2017.

vi. A screenshot of a computer program

Description automatically generated with low confidence

The last name of customer 1800 is Dickerson, they joined on Thursday.

vii. A screenshot of a computer

Description automatically generated with medium confidence

The longest gap between customer join and customer order date was 140 days during the year 2018.

B. --------------------------------------------------------------------------------------------

A screenshot of a computer

Description automatically generated with medium confidence

**C.---------------------------------------------------------------------------------------------**

**A screenshot of a computer

Description automatically generated**

**The customer ID on line 20 is 175.**

**D.--------------------------------------------------------------------------------------------**

**A screenshot of a computer

Description automatically generated**

**i. The bike with the most sales was the Revel Rail with $348,444.86.**

**ii. The Nukeproof Scout had total sales of $99,342.09.**

**E.---------------------------------------------------------------------------------------------**

**A screenshot of a computer

Description automatically generated with medium confidence**

**i. 526 different customers had returns.**

**ii. The highest amount returned by a customer was $1539.**

**iii. Something unique about this query’s results is although many customers had returns, a brief glance shows there isn’t much of a correlation between the number of items returned and the tot\_ret\_amount. Customers probably returned many different things.**

**F.---------------------------------------------------------------------------------------------**

**A screenshot of a computer

Description automatically generated**

**i. An anomaly in the query results is that the average price for all the bikes is calculated out many many decimal places.**

**ii. The total sales cost for Rocky Mountain Bikes is $782176.44.**

**G.--------------------------------------------------------------------------------------------**

**Similar to last week, the window function was a little intimidating at first. It definitely has its advantages over aggregate functions. I couldn’t imagine how long F would have taken trying to work out several lines of aggregation. It took some trial and error but after looking at it you realize it wasn’t too bad. Using partition makes dividing data into sub-categories simple. I had to get a little creative with D because it wasn’t displaying things properly, but a little searching helped me figure it out. An aggregate function could have taken much longer and taken more lines of code to get the same result. You can chop up the data more efficiently and it turns out clear in the query. Sometimes you try to aggregate over parts of data and sum totals and next thing you know nothing is working right.**