**DSCI 504: Intro to SQL**

**Assignment 02: Building the OPC Database**

**Jeffery Boczkaja**

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

i.A screenshot of a computer

Description automatically generated with medium confidence

ii.

A screenshot of a computer program

Description automatically generated with medium confidence

B. **------------------------------------------------------------------------**

A screenshot of a computer

Description automatically generated with medium confidenceA screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

1. The total sales for all hardtails is $366109.01
2. The highest selling combination is that of the DV9 and max air
3. The top NULL values just represent the total for all bikes so the names of the individual product and build names are blank. The bottom half of the table shows NULL values in the build\_name column because the data is just repeating itself.

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

A screenshot of a computer code

Description automatically generated with low confidence

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

A picture containing text, font, line, screenshot

Description automatically generated

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

A screenshot of a computer

Description automatically generated with low confidence

1. All the records for the individual who joined on 12/12/2020 were masked. A SELECT statement retrieves data from tables and it can specify the columns we want to get while an INSTERT statement is used to insert new rows of data. I believe the results were masked because the value we were using in our search criteria was a NULL value. When you search with an \*, you get all the other values in the row with the NULL value.

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

A screenshot of a computer

Description automatically generated with medium confidenceA screenshot of a computer

Description automatically generated with medium confidence A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

1. Luckily the order\_tot column was already accurate and the tax rate was accurate from before. If it wasn’t that would be okay because we could have changed the column to include the values from new\_tot!

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

A screenshot of a computer

Description automatically generated with medium confidence

1. The sum of all the phone numbers in customers from California is 395,577,183,626

H. **------------------------------------------------------------------------**

A screenshot of a computer

Description automatically generated with medium confidence A screenshot of a computer

Description automatically generated with medium confidence

1. The sum of the zip codes from customers with orders in West Virginia 4,848,445

I. **------------------------------------------------------------------------**

A screenshot of a computer

Description automatically generated

J. **------------------------------------------------------------------------**

A screenshot of a computer

Description automatically generated

K. **------------------------------------------------------------------------**

A screenshot of a computer

Description automatically generated with medium confidence

1. The cus\_app\_cd of the 500th entry is OPC83438248

L. **------------------------------------------------------------------------**

A screenshot of a computer

Description automatically generated

1. The cus\_app\_num for the 37th row is OPC85193374

M. **------------------------------------------------------------------------**

This assignment has really expanded on my data analysis skills. It was challenging and pushed me to utilize my skills. I don’t have explicit experience as a professional, but I would guess that knowing what I know now would help me simplify data for comparison. I tried to think of over-the-top ways to connect tables and sometimes the best answer is just to add another column to a table to make the translation easier. Sometimes you have to decide if you want to analyze data with a join or maybe with a cte. I now know that I can speed up the process using a cte and they aren’t as scary as they first appear. Timewise it’s not a big deal with data this size but in the real world it could make a difference when the datasets become monstrous.