STT 811

Homework 1

Due January 18, at 11:59:59 pm

All questions involve an analysis of the Northwind data. They should be answered with SQL through the sqldf command, although final answers could be found with a glimpse command used on a resultant dataframe. You do not need to output tables. Just provide code and answer questions as indicated. <u>For answers, unless otherwise indicated give answers in terms names and not just ID's</u>.

1. Perform a sort of orders by employeeID, then by shipVia, and then by freight, for those orders by shipped to France.

2. Which shipVia has the largest average cost?

3. Which product category has the highest average UnitPrice? The Lowest?

4. Which products are supplied by a company in the United States?

5. Which shipper is shipping the largest number of units of product? Answer in terms of units; you do not need to consider quantityPerUnit here.

6. Which employee is tied to the most sales revenue? Give the name, not the code, along with the total revenue for the employee.

7. Find the total revenue for each product category.

8. Consider the amount of revenue for each customer. If there were no discounts applied, which customer would see the largest increase in cost?

9. Which order(s) has the most number of items (and how many)? Give the orderID for this one.

10. Create a new field called "InventoryOrderRatio" which is, for each product, the UnitsinStock (the inventory) for the product (across all customers) divided by the quantity ordered for that product. A high value represents sufficient product in stock, while a low number represents products that are in danger of running out. What 3 products are most in danger of running out?

11. A recommender engine looks at which pairs of products tend to be bought by the same customer, so that if a customer buys one, the recommender engine will recommend they buy the other. Find which product pairs are most likely to be bought by the same customer.