

21 HW #5AO: Info Schema and Price-Volume Relationship [TBD]

First Five

Using the information schema answer the following questions:

1. Write a query which returns the count of data types (int, float, etc.) of each columns in the stocks schema.
2. Write a query which returns the number of distinct column types in the entire database.
3. Write a query which returns 3 columns: schema name, column data type, and the number of columns in that schema of that column type.
4. Rewrite the above query in a wide-format. Each row should represent a single schema.
5. Create a pie chart of the above information for the schema “information_schema”. Which data format (wide or long) did you use?

In the following exercise, we will investigate the relationship between the dollar volume of shares traded and the returns of that company. Exploring the relationship between dollar volume and return:

1. Write a query which returns the return rounded to the nearly thousandth of a percent while dealing with any data issues. Return the data in hundredths, so if the return is .037123, 3.7 should be returned. Include the dollar volume of stocks traded that day, rounded to the nearest 1,000. Also, only take a 1/16 sample using the following where statement:

```
where md5( permno::varchar(100) ) like '0%'
```

2. Create a scatter plot of your rounded returns vs. the rounded dollar volume.
3. Run simple linear regression on the rounded returns vs. the rounded dollar volume and report the results. Do you believe that there is a relationship between trading size and dollar volume traded?
4. Recreate the scatter plot making sure to remove days with less than 250 million shares traded and only include returns between -10 and 10. Did the pattern change?
5. Run simple linear regression on the rounded returns vs. the rounded dollar volume and report the results for the sample of more than 250 million shares and returns between -10 and 10. Do you believe that there is a relationship between trading size and volume traded?
6. Using only the SUM, AVG and COUNT aggregate functions, compute the variance of both the rounded volume and the rounded returns of the sample.
7. The problem below is from the analytic function lecture and should be incorporated in to the LTV estimate. Write a query which returns the following information. Cohort should be defined monthly.
 - (a) For each *complete* month, calculate the percentage revenue generated, per cohort, when compared to the previous month. For example, if Month #2 after first purchase the amount of revenue generated is equal to \$12,755.54 and the amount of money generated in Month #1 after purchase is equal to \$24,885.32 then return $\frac{12,755.54}{24,885.32} = .513376$ ¹⁶
 - (b) Average this over all the cohorts with complete month data. Be careful to only consider dates that are complete from both the start and end of the table.
 - (c) This should return a set of month-over-month multipliers that could be used to estimate the expected revenue generated from a new cohort. Explain how these numbers could be used to

¹⁶A complete month is one that is 100% in the data. For example, if a company launches on January 12, 2017 then January is not a complete month. Similarly, if today’s date is September 19th, then September is not a complete month.

estimate the lifetime value of a customer (in their first year). E.g. If a customer generated \$1 of revenue in their first month, what would you do with those multipliers to estimate the lifetime value?

DRAFT

DRAFT

Appendix D

Example Exams

This textbook is used for a variety of different courses and course formats. The exams included in this appendix reflect this diversity. In order to study from these exams, keep in mind that the amount of time given and material covered may be different.

DRAFT