# HW 9: Final Project Prep

Instructor: Jacob LaRiviere, Affiliate Professor & Director, Amazon

Email: [jlarivi1@uw.edu](mailto:jlarivi1@uw.edu)

Teaching Assistant: Lukas Hager, lghhager@uw.edu

**Due Wed Dec 7th**

Please turn in one Rmarkdown file.

Now we’re going to work with a new dataset. The goal of this is to understand customer behavior of the customers who purchase the four most high revenue items. For the final project you will have seven or eight PowerPoint slides to present to me/Lukas, your busy managers, on the business.

* This data is messy and has some weird things, like returns and shipping costs, going on with it. Be mindful of them.
* I’m not requiring it, but if you want, feel free to format the data and dates/times for extra fun analysis like tracking customers over time. Here is a resource but it isn’t needed to get 100% on the project: <http://stat.ethz.ch/R-manual/R-devel/library/base/html/as.POSIXlt.html>.

1. What percent of revenue does the top four (10?) by revenue Customer IDs account for?
2. Create a histogram of order size?
3. What is the average annual spend of each unique customer?
   1. How does it vary by country?
4. What are the top selling products and how much revenue do they account for?
5. How much variation if price is there across customers?
6. As a share of total revenue, how much revenue is returned?