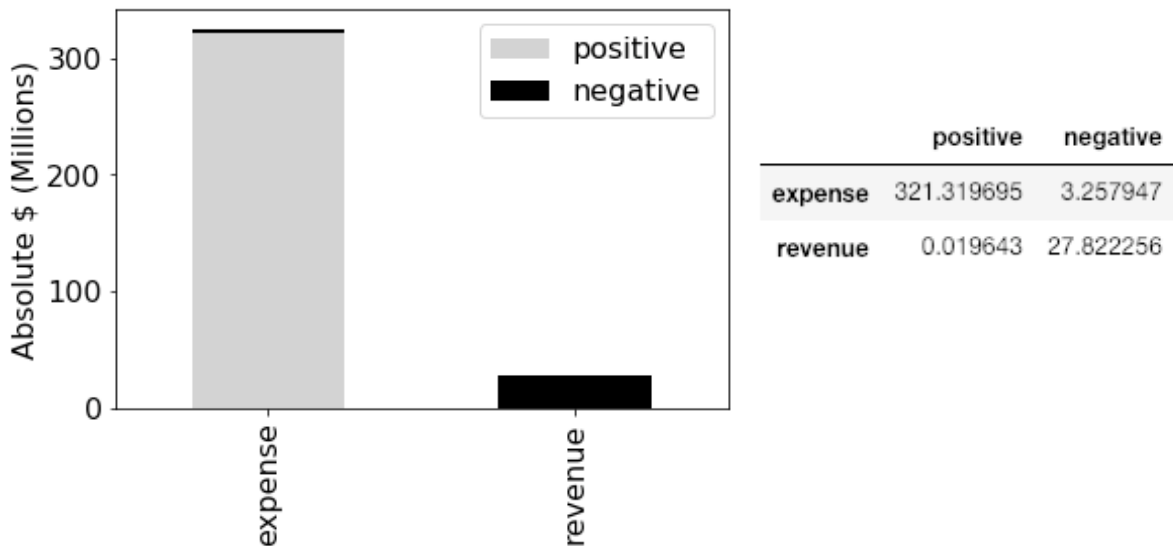
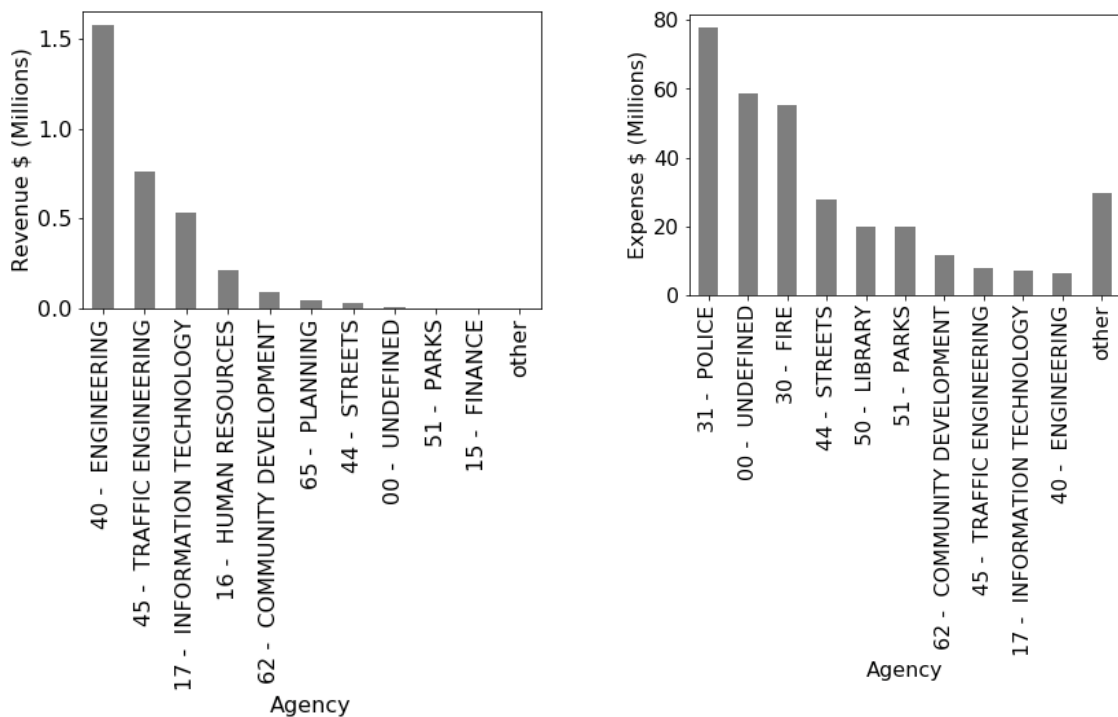


**Question 1:** How are dollars split across expense and revenue in the dataset, and how common were negative/positive values?



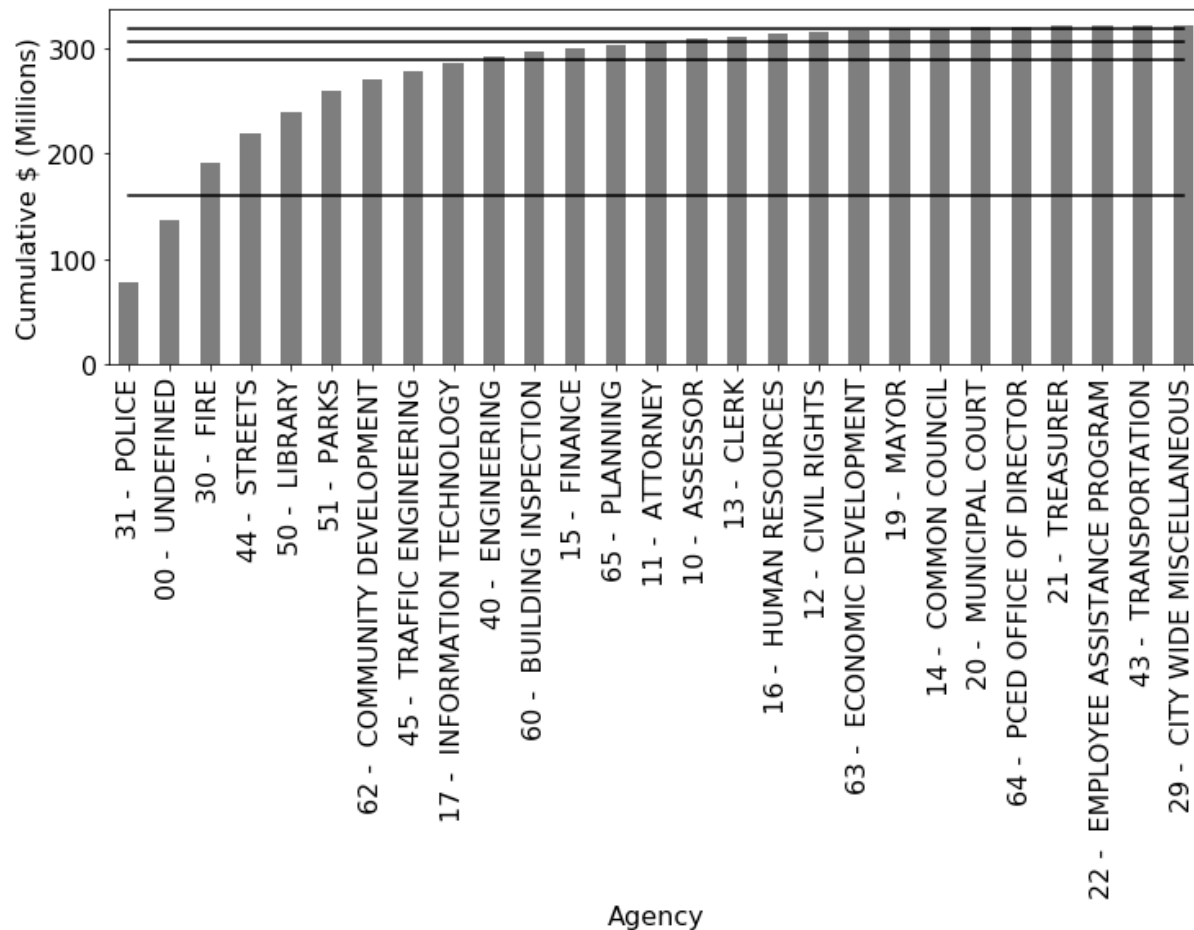
**Answer:** About 90% of data is expenses. Negative expenses are rare and positive revenues are very rare. Followup: how do you get a negative expense? Refund?

**Question 2:** what are the biggest positive expense and negative revenue organizations?



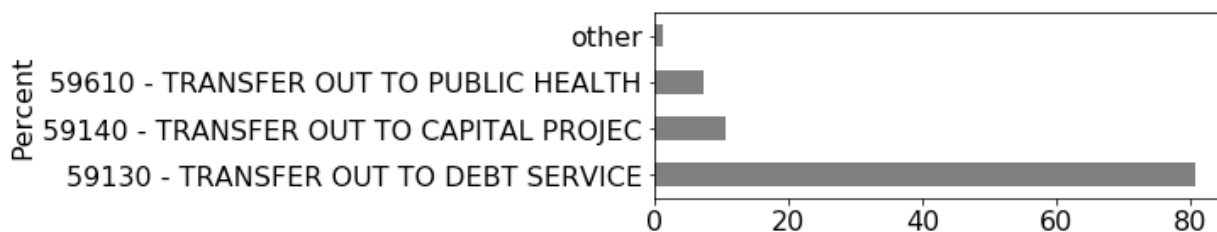
**Answer:** Engineering (both orgs) and IT dominate revenue. Police, undefined, and fire dominate expenses. The other category is significant. Followup: are revenues internal?

**Question 3:** how many agencies do we have to look at to capture 50%, 90%, 95%, 99%?



**Answer:** police, undefined, and fire get us >50%. Need 10 to get over 90%.

**Question 4:** What is undefined?



**Answer:** debt captures 80%. Capital projects and public health also count. Does forecasting undefined mostly involve predicting interest rates?

Tips:

- identify your questions and answer for each plot
- keep drilling down into large bars
- find difficult but unimportant data, and exclude it
- Use GROUP BY: `df.groupby("Agency").agg("sum")`