

Budget Study: Week 2

More git

- questions?
- merge conflicts: resolving and avoiding
- merge pull request styles:
 1. Create a merge commit
 2. Squash and merge
 3. Rebase and merge [usually best]

Choose one agency for primary focus and one for secondary focus

50	-	LIBRARY	998
45	-	TRAFFIC ENGINEERING	960
51	-	PARKS	932
31	-	POLICE	932
44	-	STREETS	923
40	-	ENGINEERING	815
30	-	FIRE	771
62	-	COMMUNITY DEVELOPMENT	491
65	-	PLANNING	274
15	-	FINANCE	254
14	-	COMMON COUNCIL	207
63	-	ECONOMIC DEVELOPMENT	150
60	-	BUILDING INSPECTION	136
16	-	HUMAN RESOURCES	112
11	-	ATTORNEY	104
17	-	INFORMATION TECHNOLOGY	101
10	-	ASSESSOR	90
13	-	CLERK	54
00	-	UNDEFINED	49
12	-	CIVIL RIGHTS	45
19	-	MAYOR	44
21	-	TREASURER	42
22	-	EMPLOYEE ASSISTANCE PROGRAM	42
43	-	TRANSPORTATION	39
20	-	MUNICIPAL COURT	39
64	-	PCED OFFICE OF DIRECTOR	38
29	-	CITY WIDE MISCELLANEOUS	1

"Even the most accurate forecast is worthless if it doesn't affect real decisions"
~ Shayne C. Kavanagh and Daniel W. Williams

Discussion

- what steps should we take before forecasting? [descriptive stats good start]
- how will we evaluate the accuracy of our forecasts? [training vs. evaluation]
- is there a way for us to "bet" on the outcomes?
- can we run some simulations and see if the MAPE (reading next week) metric has any systematic biases? Seems like it rewards under estimates...
- can we correlate against other datasets (e.g., prices on Zillow?). 80K properties and tax info here: <http://data-cityofmadison.opendata.arcgis.com/datasets/tax-parcels-assessor-property-information/data>
- conservative vs. objective forecast

Questions

- how much of the budget is captured by buckets that exist across all years? [strategy: avoid tricky data, but cap the error]
- how much data do we need to look at to capture 95% of what is happening?
- how is spending split across agencies?
- how is income split across agencies?
- how much is collected from parking tickets yearly?
- how do attorney expenses vary by year? [associate with the news?]
- how much income is from real estate taxes?
- why are some fields so low (e.g., 1, for a "PREMIUM PAY" row)
- are principal/interest payments captured here?
- how much does the City spend on forecasting?

Data cleanup

- export from Excel to CSV
- there's a mix of "-", "0", and "". They represent the same thing?
- there are spaces around the year columns
- parentheses (means negative) and commas prevent numerical ops

Cleaning up in Pandas

- `list(df)`: get list of columns in df
- `.str.replace`: do replace within each string
- `.replace`: replace whole string for exact match
- `np.nan`: may want to replace "-" with this
- `.astype(float)`: want numbers and NaN in the end