Performance Report: getTicker

Wiki versus Crunch

Data:
please see PerformanceReport_getTicker.xlsx .

Raw data is from Completedomains1010_ManualCheck.xlsx .

The Wiki and Crunch bots were applied to 214 company names ('companyName' from raw data).

The company names are under the 'company' column.
The tickers that the Wiki bot fetched are under the 'wiki' column and same for the 'crunch' column.
The 'agree?' column simply contains a bool which indicates whether or not the two bots agreed on the ticker.
The 'correct' column contains the correct ticker for the company ('manual_ticker' from raw data).
And finally, the 'who_correct' column contains the name of whichever bot fetched the correct ticker (can be 'both', 'wiki', or 'crunch').

Data Analysis:

```
In [160]: df['wiki'].value_counts().head(2)

Out[160]: PRIVATE     49
          ERR         33
          Name: wiki, dtype: int64
```

```
In [161]: df['crunch'].value_counts().head(2)

Out[161]: PRIVATE     75
          ERR          5
          Name: crunch, dtype: int64
```

```
In [162]: df['who_correct'].value_counts()

Out[162]: both       157
          wiki        36
          crunch      21
          Name: who_correct, dtype: int64
```

Overall, we can see that the Wiki bot is more successful than the Crunch bot. A total of 157 + 36 = 193 correct tickers versus 157 + 21 = 178, the difference is only of 8.1% .

However, we can notice other interesting information.

The Wiki bot seems to yield many more errors (ERR values) than the Crunch bot. This is due to the fact that Wikipedia is a resource for many things, so companies that have names that are homonymous lead

the Wiki bot to the wrong page, hence yielding an error. Also, Wikipedia has many unique or very different web pages causing the bot to sometimes be confused. Crunch, on the other hand, is a database with companies only, so whatever the bot looks for, it will find a page but not necessarily the correct page. Also, Crunch has extremely consistent web structure, meaning the bot is rarely confused as of where to fetch the information. The combination of these two factors shows why it is normal for the Crunch bot to yield a small amount of error values.

On the other hand, the Crunch bot seems to be picking up many more Private companies. I believe this is mainly due to the fact that many Wikipedia pages don't mention it when a company is private.

Conclusion

The Wiki bot seems to be more accurate (more results) but less precise (many errors) and the Crunch bot is exactly the opposite: less accurate (less results) but more precise (very little errors).

However, this is assuming that all of the web pages these bots fetched are the correct ones (meaning they fetched the web page of the company that was intended to be reached by the user). If this assumption is true, then there is no doubt that the Crunch bot is better. However, it happens more than we think that the bots get the wrong page due to issues related to the company's name or online popularity. Therefore, the Wiki bot is better because at least it warns the user whenever there is a possibility of a wrong value by yielding an error value. With the Crunch bot, it is hard to tell where the errors are because it almost always yields a result regardless.

As a result, here is how I would synthesize it:

IF you are dealing with famous companies (such as NASDAQ companies): use Crunch bot (getTickerC)

IF you are dealing with a large number of random organizations (including merchants and brands): use Wiki bot (getTickerW)

You can also use getTicker(merchant, first) if you want to use both bots (one as primary and one as backup)

Remember

If a bot is going to the wrong company web page (because another company with a similar name is more popular), you can force it to go to your select URL by doing the following:

getTickerW(getHTMLW('URL)) or getTickerC(getHTMLC('URL'))