

A1

August 21, 2022

```
[1]: import pandas as pd
      from itertools import islice
```

```
[2]: path='~/Desktop/Research Associate Assignment/Research-Aptitude-Test-Data/
      ↪Q1_Dataset.xlsx'
      df = pd.read_excel(path)
```

```
[3]: df.drop(df.columns.difference(['gvkey', 'fyear', 'csho', 'prcc_f', 'pstkl',
      ↪'lct', 'at', 'invtt', 'dltt', 'xrd', 'naics', 'sic']), 1, inplace=True)
      df.rename(columns={'at': 'att'}, inplace=True)
      df = df[df.att != 0.0]
      df = df.dropna()
      df = df.drop_duplicates()
      df = df.reset_index(drop=True)
```

```
/var/folders/6h/dmwyw_2n2078rtlzf1npp24r0000gq/T/ipykernel_13544/3614758558.py:1
: FutureWarning: In a future version of pandas all arguments of DataFrame.drop
except for the argument 'labels' will be keyword-only.
```

```
      df.drop(df.columns.difference(['gvkey', 'fyear', 'csho', 'prcc_f', 'pstkl',
      ↪'lct', 'at', 'invtt', 'dltt', 'xrd', 'naics', 'sic']), 1, inplace=True)
```

```
[4]: print(df.head())
```

	gvkey	fyear	att	csho	dltt	invtt	lct	pstkl	xrd	\
0	1072	2010	2319.482	170.142	0.0	496.495	235.665	0.0	7.392	
1	1072	2011	2468.012	169.601	0.0	566.117	297.073	0.0	7.716	
2	1072	2012	2601.995	168.633	0.0	559.074	330.877	0.0	7.150	
3	1072	2013	2384.988	168.221	0.0	550.518	161.934	0.0	10.514	
4	1072	2014	2459.015	168.190	0.0	535.912	266.309	0.0	11.951	

	prcc_f	naics	sic
0	14.91	334416	3670
1	13.26	334416	3670
2	11.90	334416	3670
3	13.18	334416	3670
4	14.27	334416	3670

```
[5]: df['tq'] = df.apply(lambda row: (((row['csho']*row['prcc_f']) + row['pstkl'] +
      ↪row['lct'] - row['att'] + row['invtt'] + row['dltt'])/row['att']), axis=1)
```

```
[6]: corr_mat = df.corr()
```

```
[7]: print("Correlation between R&D and tq is: ", round(corr_mat['xrd']['tq'], 2))
```

Correlation between R&D and tq is: -0.01

```
[8]: industries = df['naics'].unique().tolist()
fyears = df['fyear'].unique().tolist()
print(fyears)
```

[2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020]

```
[9]: cols = ['naics', 'corr']
corr_industries = pd.DataFrame(columns=cols)
for i in range(len(industries)):
    row_df = pd.DataFrame([[industries[i], round((df.loc[df.
    ↪naics==industries[i]]).corr()['xrd']['tq'], 2)]], columns=cols)
    corr_industries = pd.concat([corr_industries, row_df])
```

```
[10]: #df.loc[(df['naics'] == 334416) & (df['fyear'] == 2010)]
```

```
[11]: corr_by_time_industry = { }
for i in fyears:
    corr_by_time_industry[i] = { }
    for j in industries:
        correlation = round((df.loc[(df['fyear'] == i) & (df['naics'] == j)]).
        ↪corr()['xrd']['tq'], 2)
        corr_by_time_industry[i][j] = correlation if pd.
        ↪isna(correlation)==False else "not interpretable"
```

```
[12]: #dict(islice(corr_by_time_industry.items(), 0, 2))
print(corr_by_time_industry[2010])
```

{334416: 0.8, 334510: -0.03, 424690: 'not interpretable', 332215: -1.0, 334220: -0.03, 334413: -0.03, 333242: 0.19, 334511: -0.25, 325120: -0.83, 492110: 'not interpretable', 339112: -0.08, 325414: -0.03, 336413: 0.26, 3364: -0.72, 312140: -0.16, 337215: 'not interpretable', 334517: -1.0, 511210: -0.04, 325320: 1.0, 334513: -0.06, 332111: 'not interpretable', 424510: 'not interpretable', 213111: -0.37, 331410: -0.46, 332321: -1.0, 333132: -0.2, 31122: 'not interpretable', 445110: -0.13, 423690: 0.33, 333111: -0.52, 3252: -0.54, 334118: -0.18, 336320: -0.35, 519130: -0.04, 518210: -0.06, 322220: -0.18, 325620: -0.19, 312230: 0.2, 334514: -1.0, 325412: -0.07, 332431: 0.45, 332613: 'not interpretable', 316210: -0.4, 423830: -0.75, 517311: -0.23, 326112: 0.88, 443142: 'not interpretable', 334516: -0.26, 332216: 0.21, 336411: -0.25, 424120: 'not interpretable', 339950: 0.22, 333618: 0.79, 324110: -0.38, 541512: -0.05, 561110: 1.0, 333318: -0.19, 524114: 'not interpretable', 33441: 1.0, 333249: -0.33, 325180: -0.28, 448120: 'not interpretable', 312111: -0.66, 311422: 'not interpretable', 333316: -0.29, 324122: 'not interpretable', 331110: -0.37, 447110: 'not interpretable', 333120: -0.13, 334417: -0.39, 334290: -0.14, 722511: -0.0, 325611: -0.49, 713210: 0.14,

333413: 0.72, 325612: -0.36, 423120: 'not interpretable', 334515: 0.01, 325520: -0.53, 999990: -0.2, 334210: -0.12, 621491: 'not interpretable', 31142: 'not interpretable', 533110: -0.1, 334419: -0.22, 326211: -0.62, 339115: -0.64, 332912: -0.5, 325212: 'not interpretable', 339940: 'not interpretable', 336390: -0.19, 334112: -0.13, 452319: 'not interpretable', 531120: 0.75, 325: -0.08, 335312: -0.15, 452210: 'not interpretable', 333415: -0.31, 325199: -0.2, 452990: 'not interpretable', 333914: -0.45, 333314: 0.08, 332311: 1.0, 332510: 1.0, 335314: 0.02, 424920: 'not interpretable', 335999: -0.32, 523999: 1.0, 336340: 0.19, 326121: 'not interpretable', 621511: 0.08, 339920: -0.11, 336310: -0.4, 336120: 0.97, 325510: -0.57, 423810: 'not interpretable', 337121: -0.15, 333999: -0.18, 311812: 'not interpretable', 621512: -1.0, 33611: 'not interpretable', 423510: 'not interpretable', 325992: 'not interpretable', 448140: 'not interpretable', 336415: 'not interpretable', 999977: 1.0, 311230: -1.0, 448210: 'not interpretable', 325998: -0.49, 42: -1.0, 213112: -0.13, 313210: -0.12, 424210: -0.31, 339930: -0.01, 442110: 'not interpretable', 212221: 'not interpretable', 336412: 0.07, 335210: 'not interpretable', 3113: 'not interpretable', 33411: 'not interpretable', 325211: -0.25, 339113: -0.12, 721110: -0.1, 444110: 'not interpretable', 337214: -0.87, 336111: -0.5, 621999: -0.06, 311611: -0.26, 334418: 0.31, 33399: 1.0, 325413: -0.16, 237110: -1.0, 314110: 'not interpretable', 322130: -1.0, 452111: 'not interpretable', 333517: 1.0, 337211: 'not interpretable', 322121: 0.57, 334310: -0.54, 335122: -0.11, 311821: 'not interpretable', 423710: 'not interpretable', 337910: 1.0, 523110: 'not interpretable', 448190: 'not interpretable', 333992: 1.0, 336414: -1.0, 722514: 'not interpretable', 334519: -0.17, 333991: -1.0, 333923: -0.76, 531: 'not interpretable', 332813: 'not interpretable', 335: -0.64, 237210: 'not interpretable', 335311: 0.82, 311942: 'not interpretable', 541330: -1.0, 333994: 'not interpretable', 621: 'not interpretable', 451120: 'not interpretable', 3372: 'not interpretable', 331492: -0.85, 722513: 0.28, 424410: 'not interpretable', 311612: -1.0, 623311: 'not interpretable', 622110: 'not interpretable', 3399: 'not interpretable', 212111: 'not interpretable', 327992: 'not interpretable', 423450: 0.97, 327993: -1.0, 327213: 'not interpretable', 212: -0.14, 441310: 'not interpretable', 31211: 'not interpretable', 2111: -0.39, 442299: 'not interpretable', 531190: 'not interpretable', 335313: 0.57, 324191: 'not interpretable', 326113: -0.16, 441110: 'not interpretable', 334412: -0.14, 523930: 0.97, 423430: 0.03, 812210: 'not interpretable', 332812: 'not interpretable', 325411: -0.16, 335220: 'not interpretable', 311421: -0.44, 322219: 'not interpretable', 517312: -0.58, 323111: 0.17, 325613: 1.0, 332994: 0.61, 331210: 1.0, 3344: -1.0, 446110: 'not interpretable', 331491: 'not interpretable', 621991: -1.0, 327310: 'not interpretable', 713110: 'not interpretable', 332991: 1.0, 311520: 'not interpretable', 541380: 1.0, 333924: -0.97, 333613: 1.0, 327420: 1.0, 313110: -1.0, 3256: 'not interpretable', 211120: -0.28, 332312: 'not interpretable', 212319: 'not interpretable', 4239: 'not interpretable', 326299: -0.35, 424340: 'not interpretable', 321: 'not interpretable', 3352: -1.0, 339999: -0.09, 454110: 0.87, 336214: 'not interpretable', 448310: 'not interpretable', 335911: -0.56, 62199: 'not interpretable', 311221: 'not interpretable', 336211: 'not interpretable', 446191: 'not interpretable', 525990: 'not interpretable', 453998: 'not interpretable', 333244: 'not interpretable', 541715: 'not interpretable',

331318: 1.0, 336991: 'not interpretable', 332919: 'not interpretable', 531110:
 'not interpretable', 3363: -0.23, 325130: -1.0, 622310: 'not interpretable',
 315240: 0.92, 311615: 'not interpretable', 517911: 1.0, 326199: -0.28, 443141:
 'not interpretable', 311: 0.89, 713940: 'not interpretable', 333131: 'not
 interpretable', 522298: -0.15, 621610: 'not interpretable', 517110: 0.99,
 451130: 'not interpretable', 311511: 'not interpretable', 327120: -0.8, 523120:
 -1.0, 332999: 'not interpretable', 339114: -0.85, 4422: 'not interpretable',
 562219: 'not interpretable', 454310: 0.81, 5415: -0.29, 623110: 'not
 interpretable', 21222: 'not interpretable', 212325: -1.0, 423: 'not
 interpretable', 488190: 'not interpretable', 336999: 1.0, 721120: 'not
 interpretable', 3152: 1.0, 334111: -0.35, 311920: 'not interpretable', 453210:
 'not interpretable', 512199: -1.0, 523910: 'not interpretable', 424720: 'not
 interpretable', 332119: 'not interpretable', 333997: 'not interpretable',
 212399: 'not interpretable', 335912: -1.0, 451110: 'not interpretable', 3342:
 'not interpretable', 336360: 1.0, 335921: -1.0, 332996: 1.0, 3332: 'not
 interpretable', 561510: -1.0, 522390: 'not interpretable', 333912: 1.0, 311710:
 -1.0, 311941: 'not interpretable', 54171: 1.0, 331315: 'not interpretable',
 541519: 1.0, 423310: 'not interpretable', 3253: 0.44, 541850: 'not
 interpretable', 5311: 'not interpretable', 3359: 'not interpretable', 339994:
 'not interpretable', 454210: 'not interpretable', 515120: 'not interpretable',
 4481: 'not interpretable', 312120: -1.0, 541810: 'not interpretable', 32419:
 'not interpretable', 454111: -0.25, 335932: 'not interpretable', 315280: 'not
 interpretable', 5418: -1.0, 611310: 'not interpretable', 3212: 'not
 interpretable', 611420: 'not interpretable', 3351: 'not interpretable', 561520:
 'not interpretable', 333611: -0.95, 488320: 'not interpretable', 115114: 'not
 interpretable', 621493: 'not interpretable', 321911: 'not interpretable', 3222:
 1.0, 517919: -0.28, 237130: -0.98, 311412: 'not interpretable', 2122: 'not
 interpretable', 311111: 'not interpretable', 423490: 'not interpretable',
 337122: 'not interpretable', 32519: 'not interpretable', 444190: 'not
 interpretable', 313320: 'not interpretable', 423840: 'not interpretable',
 452910: 'not interpretable', 4233: 'not interpretable', 33299: 'not
 interpretable', 112511: 'not interpretable', 112512: 'not interpretable',
 451211: 'not interpretable', 332993: 'not interpretable', 4231: 'not
 interpretable', 541360: -0.67, 321999: 'not interpretable', 452311: 'not
 interpretable', 515111: 'not interpretable', 444220: 'not interpretable',
 326212: 'not interpretable', 212291: -1.0, 562211: -1.0, 562213: 'not
 interpretable', 336212: 'not interpretable', 311911: 'not interpretable',
 336612: 'not interpretable', 221118: 'not interpretable', 713290: 'not
 interpretable', 448110: 'not interpretable', 621340: 'not interpretable',
 325314: 'not interpretable', 322212: 'not interpretable', 424490: 'not
 interpretable', 331420: -1.0, 541620: 'not interpretable', 423820: 'not
 interpretable', 335991: 1.0, 333243: -0.86, 332410: 0.83, 441210: 'not
 interpretable', 325194: 'not interpretable', 3251: 'not interpretable', 7132:
 'not interpretable', 611630: 'not interpretable', 311340: 'not interpretable',
 3313: 'not interpretable', 333414: 'not interpretable', 325220: 'not
 interpretable', 212230: 'not interpretable', 238910: 'not interpretable',
 327212: 'not interpretable', 333241: 1.0, 453910: 'not interpretable', 713990:
 'not interpretable', 236115: 'not interpretable', 4244: 'not interpretable',

311411: 'not interpretable', 517410: -0.33, 316998: 'not interpretable', 424940:
 1.0, 331221: 'not interpretable', 441222: 'not interpretable', 326140: 'not
 interpretable', 311999: -1.0, 561311: -0.16, 512110: 'not interpretable',
 211111: 'not interpretable', 425120: 'not interpretable', 315220: 'not
 interpretable', 336350: 1.0, 325193: 'not interpretable', 111339: 'not
 interpretable', 532412: 'not interpretable', 515112: 'not interpretable',
 237120: -1.0, 623990: 'not interpretable', 212299: 'not interpretable', 523210:
 'not interpretable', 3339: 1.0, 4483: 'not interpretable', 531210: 'not
 interpretable', 522291: 'not interpretable', 541511: 0.94, 446130: 'not
 interpretable', 4539: 'not interpretable', 333612: 'not interpretable', 423520:
 'not interpretable', 611699: -1.0, 45411: 'not interpretable', 611: 'not
 interpretable', 541890: 1.0, 4248: 'not interpretable', 3114: 'not
 interpretable', 485310: 'not interpretable', 111: 1.0, 61: 'not interpretable',
 4411: 'not interpretable', 44122: 'not interpretable', 524210: 'not
 interpretable', 621498: 'not interpretable', 611710: 'not interpretable',
 488999: 'not interpretable', 519: 'not interpretable', 333921: 'not
 interpretable', 32191: 'not interpretable', 454: 'not interpretable', 561499:
 -1.0, 335129: 'not interpretable', 541713: 'not interpretable', 721: 'not
 interpretable', 515210: 'not interpretable', 722: 'not interpretable', 336510:
 1.0, 561312: 'not interpretable', 621111: 'not interpretable', 722410: 'not
 interpretable', 423910: 'not interpretable', 621492: 1.0, 561990: 'not
 interpretable', 327390: 'not interpretable', 623312: 'not interpretable',
 322110: 'not interpretable', 541712: 'not interpretable', 722515: 'not
 interpretable', 541219: 'not interpretable', 221114: 'not interpretable',
 423610: 'not interpretable', 327999: 0.91, 48111: 'not interpretable', 561611:
 'not interpretable', 332911: -0.9, 448150: 'not interpretable', 711212: 'not
 interpretable', 541714: -1.0, 335110: 1.0, 481111: 'not interpretable', 441120:
 'not interpretable', 335929: 'not interpretable', 522320: 1.0, 562: 'not
 interpretable', 315: -0.21, 3314: 'not interpretable', 311119: 'not
 interpretable', 336330: -1.0, 448130: 'not interpretable', 3369: 'not
 interpretable', 315990: 'not interpretable', 336112: 'not interpretable',
 621210: 'not interpretable', 311314: 'not interpretable', 2211: 1.0, 511130:
 'not interpretable', 5621: 'not interpretable', 2123: 'not interpretable',
 321113: 'not interpretable', 221115: -0.96, 446120: 'not interpretable', 2121:
 -1.0, 311225: 'not interpretable', 333996: 'not interpretable', 31521: 'not
 interpretable', 424820: 'not interpretable', 423930: 'not interpretable',
 524291: 'not interpretable', 321219: 'not interpretable', 5222: 'not
 interpretable', 212393: 'not interpretable', 541611: 0.95, 541910: 'not
 interpretable', 423990: 'not interpretable', 311514: 1.0, 322211: 'not
 interpretable', 561450: 'not interpretable', 447190: 'not interpretable',
 488330: 'not interpretable', 311930: 'not interpretable', 486: 'not
 interpretable', 237: 'not interpretable', 561320: 'not interpretable', 311211:
 -1.0, 336992: 'not interpretable', 331313: 'not interpretable', 532282: 'not
 interpretable', 424480: 'not interpretable', 339991: 'not interpretable',
 622210: 'not interpretable', 4238: 'not interpretable', 423140: 'not
 interpretable', 512120: 'not interpretable', 561621: 'not interpretable',
 425110: 'not interpretable', 111419: 'not interpretable', 111998: 'not
 interpretable', 523130: 'not interpretable', 33591: 'not interpretable', 524126:

'not interpretable', 32599: 'not interpretable', 522292: 'not interpretable', 32311: 'not interpretable', 3323: 'not interpretable', 454113: 'not interpretable', 445310: 'not interpretable', 444130: 'not interpretable', 327991: 'not interpretable', 326122: 'not interpretable', 531390: 'not interpretable', 323117: 'not interpretable', 311224: 'not interpretable', 812220: 'not interpretable', 532111: 'not interpretable', 611610: 'not interpretable', 483111: 'not interpretable', 541711: 'not interpretable', 423110: 'not interpretable', 532411: 'not interpretable', 324199: 'not interpretable', 441228: 'not interpretable', 611110: 1.0, 6214: 'not interpretable', 522310: 'not interpretable', 212391: 'not interpretable', 332: 'not interpretable', 3261: 'not interpretable', 524298: 'not interpretable', 424710: 'not interpretable', 541840: 'not interpretable', 541613: 'not interpretable', 33991: 'not interpretable', 321114: 'not interpretable', 238220: 'not interpretable', 523920: 'not interpretable', 236220: 'not interpretable', 112210: 'not interpretable', 33712: 'not interpretable', 33142: 'not interpretable', 3311: 'not interpretable', 561422: 'not interpretable', 111219: 'not interpretable', 336611: 'not interpretable', 722320: 'not interpretable', 524292: 'not interpretable', 33361: 'not interpretable', 3361: 'not interpretable', 1114: 'not interpretable', 326150: 'not interpretable', 337110: 'not interpretable', 621420: 'not interpretable', 711211: 'not interpretable', 325110: 'not interpretable', 212210: 'not interpretable', 312130: 'not interpretable', 517210: 'not interpretable', 2361: 'not interpretable', 4237: 'not interpretable', 3259: 'not interpretable'}