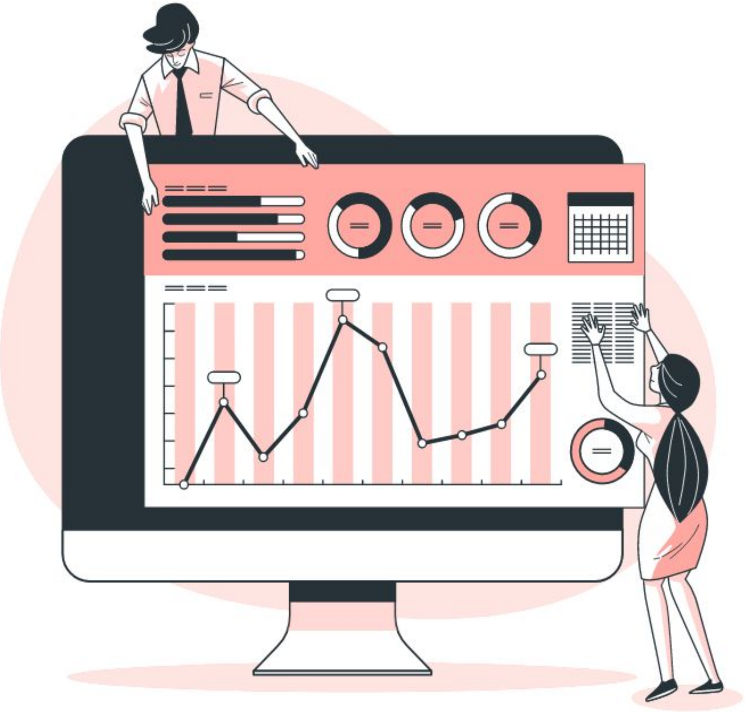
BA870
Final Project

Yilun Wang:
yilun830@bu.edu



Risk Exposures

Data: Tickers and RetYTD have 1886 values.

RetYTD are objects. I converted it to floats.

```
data_return['RET'] = pd.to_numeric(data_return['RET'],errors='coerce')
```

After converting, RetYTD have 86 null values. I fill them with 0.

```
# check NA again
data_return.isna().sum()

PERMNO     0
date       0
TICKER    62
RET        0
dtype: int64
```

Ignored missing values in TICKER. No null values in FF factors dataset.

Risk Exposures

- R-squared: 0.095
- Adj.R-squared: 0.093
- Mktrf's p-value bigger than 0.05
- Coefficient of risk exposure:

  - Market : 0.0082

  -Size : -0.0120

  -Value : 0.1068

OLS Regression Results

Dep. Variable:	RetYTD	**R-squared:**	0.095
Model:	OLS	**Adj. R-squared:**	0.093
Method:	Least Squares	**F-statistic:**	65.51
Date:	Thu, 28 Apr 2022	**Prob (F-statistic):**	2.76e-40
Time:	01:50:54	**Log-Likelihood:**	79.945
No. Observations:	1886	**AIC:**	-151.9
Df Residuals:	1882	**BIC:**	-129.7
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	-0.0981	0.011	-8.783	0.000	-0.120	-0.076
mktrf	0.0082	0.009	0.957	0.339	-0.009	0.025
smb	-0.0120	0.004	-2.736	0.006	-0.021	-0.003
hml	0.1068	0.008	13.814	0.000	0.092	0.122

Omnibus:	642.171	**Durbin-Watson:**	2.005
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3599.566
Skew:	1.492	**Prob(JB):**	0.00
Kurtosis:	9.074	**Cond. No.**	4.56

Financial ratios

Ratios:

```
23   book/price
24   e/price
25   sale/price
26   ebit/price
27   price/cashflow
28   totaldebt/totalassets
29   totaldebt/equity
30   roa
31   totaldebt/totalliabilies
32   totaldebt/capital
```

Data: 1886 values

Missing values in :
-price/cashflow: 1
-totaldebt/totalassets: 237
-totaldebt/equity: 5
-totaldebt/totalliab: 5
-totaldebt/capital: 237

Winsorize + average value

Financial ratios

- R-squared: 0.085
- Adj.R-squared: 0.080
- Coefficient: 6 positive + 4 negative
- 4 of P-values bigger than 0.05

Dep. Variable:	RetYTD	R-squared:	0.085
Model:	OLS	Adj. R-squared:	0.080
Method:	Least Squares	F-statistic:	17.46
Date:	Thu, 28 Apr 2022	Prob (F-statistic):	9.87e-31
Time:	01:50:54	Log-Likelihood:	70.268
No. Observations:	1886	AIC:	-118.5
Df Residuals:	1875	BIC:	-57.57
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.1639	0.017	-9.731	0.000	-0.197	-0.131
book/price	0.1370	0.014	9.704	0.000	0.109	0.165
e/price	-0.1240	0.063	-1.965	0.050	-0.248	-0.000
sale/price	0.0130	0.004	3.251	0.001	0.005	0.021
ebit/price	0.1436	0.071	2.020	0.043	0.004	0.283
price/cashflow	-0.0001	0.000	-1.105	0.269	-0.000	9.49e-05
totaldebt/totalassets	0.0759	0.041	1.855	0.064	-0.004	0.156
totaldebt/equity	-0.0020	0.001	-1.976	0.048	-0.004	-1.49e-05
roa	0.1464	0.044	3.322	0.001	0.060	0.233
totaldebt/totalliabilies	-0.0155	0.033	-0.463	0.644	-0.081	0.050
totaldebt/capital	0.0277	0.026	1.078	0.281	-0.023	0.078

Omnibus:	649.740	Durbin-Watson:	1.948
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4325.518
Skew:	1.454	Prob(JB):	0.00
Kurtosis:	9.826	Cond. No.	849.

Industries

Data:

No null value.

After get dummies, 25 columns.

```
0   Ticker        1886 non-null   object
1   ggroup_1010   1886 non-null   uint8
2   ggroup_1510   1886 non-null   uint8
3   ggroup_2010   1886 non-null   uint8
4   ggroup_2020   1886 non-null   uint8
5   ggroup_2030   1886 non-null   uint8
6   ggroup_2510   1886 non-null   uint8
7   ggroup_2520   1886 non-null   uint8
8   ggroup_2530   1886 non-null   uint8
9   ggroup_2550   1886 non-null   uint8
10  ggroup_3010   1886 non-null   uint8
11  ggroup_3020   1886 non-null   uint8
12  ggroup_3030   1886 non-null   uint8
13  ggroup_3510   1886 non-null   uint8
14  ggroup_3520   1886 non-null   uint8
15  ggroup_4010   1886 non-null   uint8
16  ggroup_4020   1886 non-null   uint8
17  ggroup_4030   1886 non-null   uint8
18  ggroup_4510   1886 non-null   uint8
19  ggroup_4520   1886 non-null   uint8
20  ggroup_4530   1886 non-null   uint8
21  ggroup_5010   1886 non-null   uint8
22  ggroup_5020   1886 non-null   uint8
23  ggroup_5510   1886 non-null   uint8
24  ggroup_6010   1886 non-null   uint8
```

Industries

- R-squared: 0.326

- Adj.R-squared: 0.319

- Coefficient:

  9 positive + 15 negative

- 7 of P-values bigger than 0.05

OLS Regression Results

Dep. Variable:	RetYTD	R-squared:	0.326
Model:	OLS	Adj. R-squared:	0.317
Method:	Least Squares	F-statistic:	39.07
Date:	Thu, 28 Apr 2022	Prob (F-statistic):	6.29e-141
Time:	01:50:54	Log-Likelihood:	357.63
No. Observations:	1886	AIC:	-667.3
Df Residuals:	1862	BIC:	-534.2
Df Model:	23		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0525	0.006	-9.154	0.000	-0.064	-0.041
ggroup_1010	0.5675	0.023	24.634	0.000	0.522	0.613
ggroup_1510	0.1103	0.021	5.346	0.000	0.070	0.151
ggroup_2010	-0.0381	0.015	-2.473	0.013	-0.068	-0.008
ggroup_2020	0.0173	0.025	0.692	0.489	-0.032	0.066
ggroup_2030	-0.0150	0.031	-0.476	0.634	-0.077	0.047
ggroup_2510	-0.1555	0.039	-3.980	0.000	-0.232	-0.079
ggroup_2520	-0.1561	0.025	-6.242	0.000	-0.205	-0.107
ggroup_2530	-0.0101	0.025	-0.407	0.684	-0.059	0.039
ggroup_2550	-0.0987	0.022	-4.569	0.000	-0.141	-0.056
ggroup_3010	0.1308	0.047	2.770	0.006	0.038	0.223
ggroup_3020	0.0913	0.030	3.074	0.002	0.033	0.149
ggroup_3030	-0.1013	0.047	-2.145	0.032	-0.194	-0.009
ggroup_3510	-0.0140	0.018	-0.755	0.450	-0.050	0.022
ggroup_3520	-0.1115	0.016	-6.887	0.000	-0.143	-0.080
ggroup_4010	-0.0354	0.016	-2.275	0.023	-0.066	-0.005
ggroup_4020	-0.0700	0.021	-3.344	0.001	-0.111	-0.029
ggroup_4030	0.0563	0.025	2.215	0.027	0.006	0.106
ggroup_4510	-0.0562	0.019	-3.005	0.003	-0.093	-0.020
ggroup_4520	-0.1196	0.022	-5.412	0.000	-0.163	-0.076
ggroup_4530	-0.2003	0.027	-7.509	0.000	-0.253	-0.148
ggroup_5010	0.0643	0.054	1.193	0.233	-0.041	0.170
ggroup_5020	-0.0184	0.030	-0.606	0.545	-0.078	0.041
ggroup_5510	0.0932	0.026	3.584	0.000	0.042	0.144
ggroup_6010	0.0167	0.018	0.934	0.350	-0.018	0.052

Combination

After combination:

- R-squared : 0.0382
- Adj.R-squared : 0.0370

Number of positive and negative coefficients did not change too much.

Number of P-values which are bigger than 0.05 increased.

```
# Run OLS regression
y = dfff["RetYTD"]
X = dfff[['mktrf','smb','hml','book/price','e/price','sale/price','ebit/price','price/cashflow','
          'roa','totaldebt/totalliabilities','totaldebt/capital','ggroup_1010','ggroup_1510',
          'ggroup_2010','ggroup_2020','ggroup_2030','ggroup_2510','ggroup_2520','ggroup_2530','gg
          'ggroup_3010','ggroup_3020','ggroup_3030','ggroup_3510','ggroup_3520',
          'ggroup_4010','ggroup_4020','ggroup_4030','ggroup_4510','ggroup_4520','ggroup_4530','gg
          'ggroup_5020','ggroup_5510','ggroup_6010']]
# Use statsmodels
X = sm.add_constant(X) # adding a constant
model = sm.OLS(y, X).fit()
model.summary()
```

Dep. Variable:	RetYTD	R-squared:	0.382
Model:	OLS	Adj. R-squared:	0.370
Method:	Least Squares	F-statistic:	31.71
Date:	Thu, 28 Apr 2022	Prob (F-statistic):	5.93e-165
Time:	01:50:54	Log-Likelihood:	439.74
No. Observations:	1886	AIC:	-805.5
Df Residuals:	1849	BIC:	-600.4
Df Model:	36		

End

THANK YOU!!