**Introduction**: There are several attempts made to analyze the factors which impact asset (stock) return. One of the earliest models was Capital Asset Pricing Model (CAPM) (1,2) and ICAPM (5), which explain how a stock performs compared to overall market. CAPM’s simplicity leaves a large chunk of stock performance unexplained, which led Fama and French (FF) to propose a 3 factor asset pricing model in 1993 (3). The theory continuously evolved to a proposal that the factors are just diversified portfolios that provide different combinations of exposures to the unknown state variables (FF 1996) (6) and then an augmented 5-factor model in 2015 (4)

The FF five factor model sought to explain a stock’s price movement with the following determinants.

1. **Market Return**: – is the overall market performance.
2. **SMB** (**S**mall **M**inus **B**ig) is the return advantage / disadvantage of small firms’ size over big firms’ size measured by firm’s market capitalization (price times shares outstanding)
3. **HML** (**H**igh value **M**inus **L**ow Value –book value to market value ratio) is the return advantage / disadvantage of cheaply priced firms versus expensive firms.
4. **RMW (R**obust **M**inus **W**eak) is the difference between highly profitable companies against less profitable ones.
5. **CMA****(C**onservative **M**inus **A**ggressive) is the return advantage / disadvantage of companies with conservative investment decisions compared to companies in expansion.

This model is extremely popular. Factor investing strategy is followed by large investment companies. Multiple research papers followed this paper for validating model effectiveness and also trying to expand the theory. (7, 8, 9, 10,11, 12, 13, 14, 15). There is also another factor “Momentum” (16) which is used as a factor investing strategy. If the stock momentum crosses market momentum, then that stock is likely to outperform market. Fama and French model does not explicitly consider momentum factor.

Our objective is to develop a data pipeline and model to visualize and validate Fama and French five factor asset pricing model. We will use free time-series stock and fundamental data APIs as our data source for this analysis. Project team will analyze NYSE scripts. We will build the five-factor model on the data using Python/R/MySQL and create an interactive visualization (using D3 / JavaScript) and use Google Cloud Service for hosting. We will validate model hypothesis.

We want to build a standardized algorithm pipeline which can be plugged to any data source (any stock market across the globe). Using our algorithm and interactive platform, we will offer investors who would like to explore factor investing a comprehensive picture to stock selection. With the detailed breakdown of individual stock's factor contribution and sensitivity, everyone could construct their own "smart-beta" portfolio just like established fund managers do.

**How is it done today; what are the limits of current practice?**

Big investment firms use Factor Investing strategy. In the USA, there are many funds which focus on a specific factor or combination of factors. For example, Size fund, Momentum fund, smart beta fund etc. We are not seeing large scale use of factor investing in non-USA markets. May be factor investing is a new concept for other markets.

**What is new in our approach? Why will it be successful?**

We will create a toolkit which will help break technical barrier for small investors. We think that by making it easier for investors to invest based on factors will be an excellent value proposition for many personal, small, and medium investors.

**Who cares?**

Many personal, small, and medium investors from the USA and non-USA markets will benefit from this information.

**If we are successful, what difference and impact will it make, and how do we measure them?**

Like a Play/App store App, success for our service is measured by number of users using this service and average feedback ratings.

**What are the risks and payoffs?**

As market is self-correcting, if more investors shift to factor investing, the overall returns may diminish over time. Once these factors become less effective, we may have to find other factors affecting stock price movements.

**How much will it cost?**

To start with we will use free to use APIs (which have a limitation on calls per minute or calls per day). But this is not an issue for academic project. We will use open-source technology to build the prototype – hence we do not expect any major cost developing this project.

But if we make this service commercial then we will have to buy API subscriptions (subscription costs defer from market to market and provider to provider), cloud hosting charges, security certification charges etc.

**How long will it take?**

The project team is confident to build the prototype consisting of end-to-end data pipeline + model algorithm including interactive visualizations within a period of 8 to 10 weeks (about 2.5 months).

**What are the midterm and final “exams” to check for success? How will progress be measured?**

The project team is a mix of functional experts, project management professionals and full stack developers. As a team, we have decided following milestones and will divide the work among the team members based on their area of expertise.

1. Design specifications including user stories, functional solution, and technical solution specification. This will include data source, data structure, data transformation, data flow diagrams, data model – entity relationships, data analysis algorithms, user experience wireframes etc. - 4 weeks.
2. Develop, test, refine cycle(s) – 3 weeks.
3. Final updated design documentation– 1 week
4. Final presentation – 2 weeks

Progress will be measured using stage gate checklists for each of the stages above.

All team members have contributed similar amount of effort so far in following activities:

* exploring project problem statements
* conducting literature survey
* identifying possible data sources,
* discussing data transformation and high level design
* project deliverables planning and coordination

**Project Roles**

* Project Manager - Normando Bernardo
* Technical Experts - Shivani Bhor and Prakash Selvam
* Functional Experts - Yi-Hsuan, Lu and Rishikesh Gaikwad

**Proposed method (should be almost finished)**

We are going use free market APIs (Alphavantage ) for stock data from NYSE exchange. APIs are available for both stock ticks and fundamental data of NYSE companies. Using fundamental data API, we can find out company size, value, investment and profit parameters. Using these fundamental parameters, we will classify each stock as Small/Big size, Low / High Value, Investing/ Paying owners, Profitable / Not profitable etc. Each of these pairs when combined with other pairs will help in classifying the stocks in 16 different categories. Each category is type of a portfolio. Once the stock portfolios are defined, then we need to compare the performance of these combination portfolios against market over time. For this comparison, we will use stock tick data ( open, close, high , low ) over a period of time. We can understand how the stock portfolio has performed compared to market and currency.

We will do a regression analysis to understand which factors had a larger correlation to the stock price. We will also validate multicollinearity, ANOVA and see if any transformation is required to arrive at a better fit.

Once the model is established, then we will dive further into visualization creation. We plan to have an user interface where a small investor can select a stock code, and see how the stock is performing against market and also see the stock classification ( SMB, HML etc,,). We can also list the stocks falling in a particular portfolio in a table view and show the current performance in terms of beta at each stock level and also at portfolio level.

Investor can then make investment decision based on his or her risk appetite and with knowledge of how a specific portfolio return compares to market returns.

**Design of upcoming experiments / evaluation**

1. **Regression Analysis and Fama and French model validation**

**Plan of activities (please show the old one and the revised one, along with the activities of each group member)**

1. **Data download and cleaning – Shivani and Prakash**
2. **Stock classification and time series data – All team members**
3. **Progress Report – Norman supported by all team members.**
4. **Data pipeline and visualization development – Functional and Technical team**

**Clear**list of innovations**: give a list of the best 2-4  ideas that your approach exhibits.**

Use regression analysis , ANOVA and test model fit , choose appropriate transformations to arrive a best fit prediction model.

**References:**

1. Sharpe, William F., 1964. Capital asset prices : A theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.
2. Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.
3. Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.
4. Fama, E., French, K., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.
5. Merton, R., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887.
6. Fama, E., 1996. Multi factor portfolio efficiency and multifactor asset pricing. Journal of Financial and Quantitative Analysis 31, 441–465.
7. James Foye, 2018, A comprehensive test of the Fama-French five-factor model in emerging markets, Emerging Markets Review 37, 199-222
8. Wilma de Groota, Joop Huij, 2018, Are the Fama-French factors really compensation for distress risk?, Journal of International Money and Finance, 86, 50-69
9. María de la O González, Francisco Jareño, 2019, Testing extensions of Fama & French models: A quantile regression approach, The Quarterly Review of Economics and Finance, 71, 188-204
10. Hao Liu , Ya-Chun Gao, 2019, The impact of corporate lifecycle on Fama–French three-factor model, Physica A: Statistical Mechanics and its Applications, 513, 390-398
11. Jan Bartholdy, Paula Peare, 2005, Estimation of expected return: CAPM vs. Fama and French, International Review of Financial Analysis 14-4, 407-427
12. Po-Chin Wu, Shiao-Yen Liu, Che-Ying Chen, 2016, Re-examining risk premiums in the Fama–French model: The role of investor sentiment, The North American Journal of Economics and Finance, Volume 36Pages 154-171
13. Marshall L. Stocker, 2016, The price of freedom: A Fama–French freedom factor, Emerging Markets Review 26, 1-19
14. Eugene F. Fama, Kenneth R. French, 2018, Choosing factors, Journal of Financial Economics 28 February 2018, Volume 128, Issue 2 (Cover date: May 2018) Pages 234-252
15. Eugene F. Fama, Kenneth R. French, 2016, International tests of a five-factor asset pricing model, Journal of Financial Economics, 23 November 2016 Volume 123, Issue 3 (Cover date: March 2017) Pages 441-463
16. Chu-Sheng Tai, 2003, Are Fama–French and momentum factors really priced?, Journal of Multinational Financial Management , December 2003 , Volume 13, Issues 4–5 , Pages 359-38