Study of Social Enterprises Using Data Science

Original - 13th June’21; Modified - 15th September’21

By - Pranjita Chakraborty

External Supervisor: Prof. Amelie Artis (Science Po. Grenoble, France)
Academic Mentor and Supervisor: Prof. Dmitry I. Ignatov. (School of DAAI, HSE, Moscow)

Scope of Discussion

- Interpreting financial models,
- analysing alternate funding mechanism - crowdfunding,
- marketing plan using network metrics

Problem Statement and Economic significance

Social enterprises - especially nonprofits saw a decline in grants and contributions; considered not making payroll; went out of business.

Social enterprise being an emerging field, lack of enough research literature in this area.

According to the CIRIEC report 2017, about 6.3% of the EU workforce (or 13.6 million jobs) were employed in this sector. Hence unemployment in this sector hits the economy severely.

Step - 1: Current research trends

Results of Topic Modeling

RA K E	LDA
'social enterprises serving','organizational form relates', 'entrepreneurial projects requires', 'dominant organizational form', 'address practical issues',, 'providing financial services', 'financial markets', 'simple framework', 'resource acquisition', 'relevant literature',, 'performance metrics', 'manuscript reviews', 'main groups', 'long tradition', 'limited access', 'foundational logic', 'emerging markets', 'costly resources', 'capital structure', 'nonprofit mfis', 'funding sources',	1. Social, 2. Public, 3. Base, 4. Government, 5. Service, 6. Cost, 7. Level, 8. Volunteer, 9. Nonprofit, 10. Study, 11. Performance, 12. Financial, 13. Sector, 14. Community, 15. Hospital, 16. Profit, 17. Effect, 18. Stakeholder, 19. Organization, 20. Market, 21. Impact, 22. Model, 23. Value, 24. Work, 25. Identify, 26. Article, 27. Donation, 28. Business, 29. Process, 30. Donor.

Results of Topic Modeling

Conceptualised Topic Model

1. Likes, 2. De, 3. Favorable, 4. Maximize, 5. Managerialism, 6. Person, 7. Pricing, 8. Cash, 9. Resilience, 10. Assesses, 11. Transaction, 12. Narratives, 13. Uncertainty, 14. Dialogic, 15. Canadian, 16. Distinctive, 17. Partial, 18. Police, 19. Contingent, 20. Mix, 21. Class, 22. Security, 23. Italy, 24. Meet, 25. Enables, 26. Responsibilities, 27. Score, 28. Squares, 29. Allow, 30. Article.

BERTopic

1. Nonprofit organizational management data
2. Environment entrepreneurship

Results of Topic Modeling

NMF

- social, enterprises, media, innovation, impact, profit, community, economic, business, communication
- leadership, board, organizational, governance, members, leaders, nonprofit, organizations, periodicals,
- donors, donations, giving, charity, charitable, effect, trust, study, marketing
- volunteers, engagement, service, management, organization, quality, time
-npos, accountability, funding, business, financial, practices, context, stakeholders
-performance, profit, use, information, data, nonprofit

Step 2. - Interpret the financial model

Important Features denoted by Feature selection

current assets (like cash, cash equivalents, accounts receivable, stock inventory, marketable securities, prepaid liabilities, and other liquid assets which are either to be sold or will be used for conducting short-term business operations);

current liabilities (which are short term financial obligations due within one business cycle (usually one year))

operating income (which is the difference between operating revenue and operating expenses) are of utmost importance.

Thus, we infer that the financial model of these enterprises should be based upon short term financial analysis focussed upon operations turnover and liquidity

The data collected for step - 2

Internal Data:

1. Provided by Laboratory - total 24 observations (Used 23 for study as Fusees had very few variables)
2. French enterprises: 18 nonprofits, 3 cooperatives, 3 unions
3. Sectors: a) 9 - Social Activities
   b) 2 - Art and culture, cycling & awareness, Employment
   c) one each - Integration through economic activity, Heritage reconstruction, Maintaining green spaces; Tourism; education; entrepreneurship; ecology; funding; environment; union.
4. After adding ratios - total 40 columns (34 financial variables)

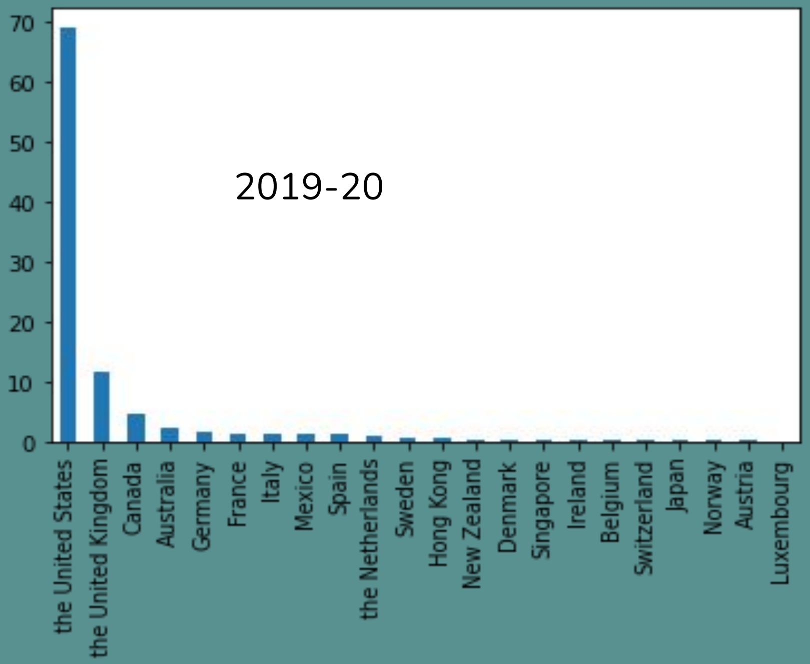
External Data:

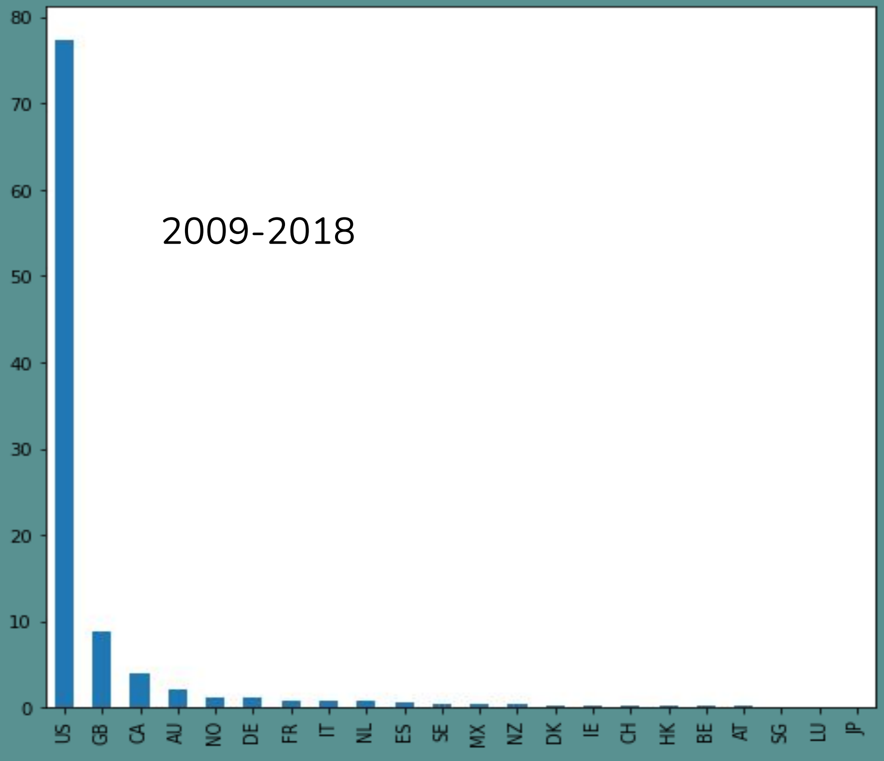
1. Financial and other Data collected from the Tax files of nonprofits with Internal Revenue Services
2. Contains only nonprofits based in US.
3. Used only observations from California to avoid any representation bias from the range of number of nonprofits per state (2 to 30604)
4. Individual sectoral information wasn’t available
5. To be queried using BigQuery from google cloud storage

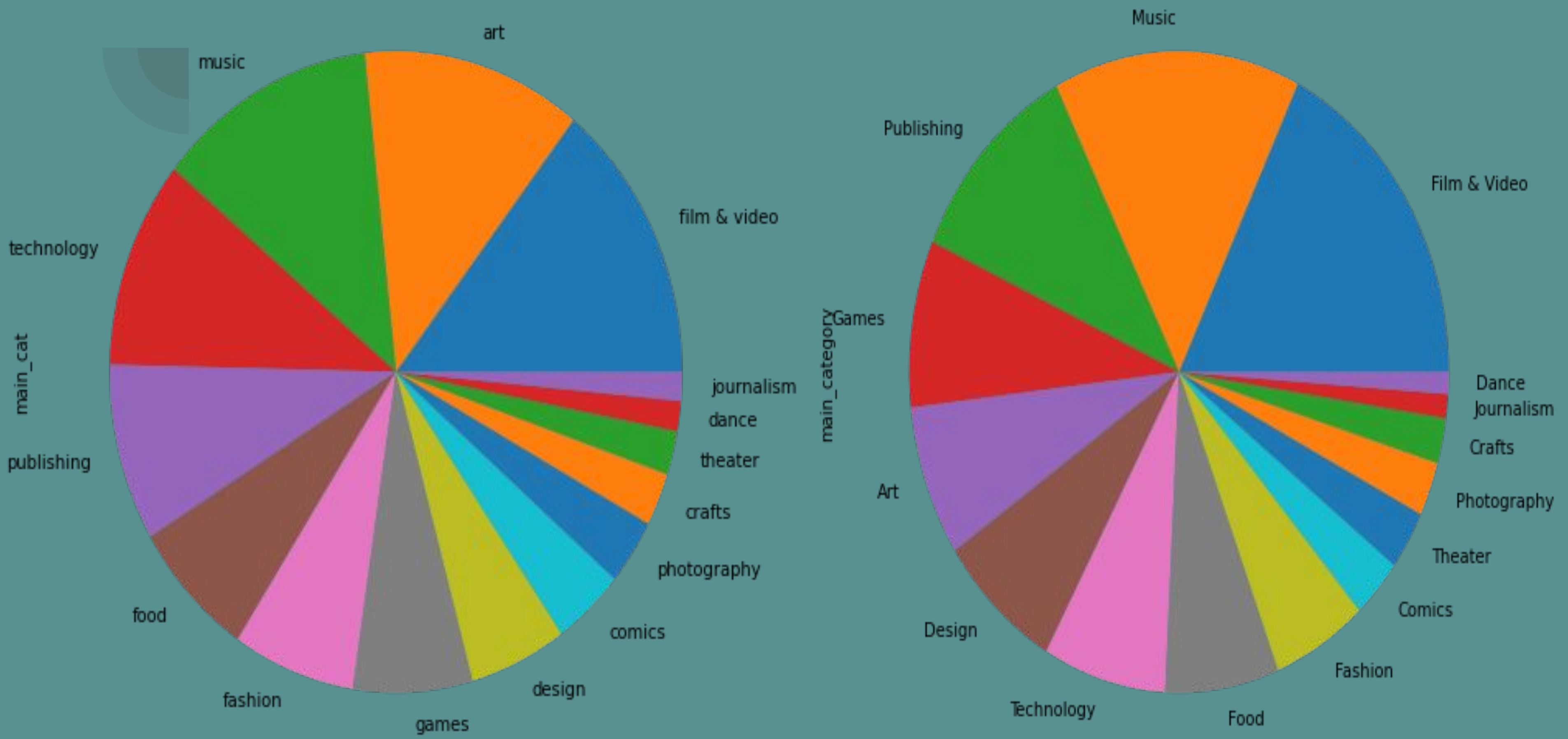
Step - 3: Analyzing alternate funding source - crowdfunding

Backdrop:
1) Search for alternative streams of funding beyond government donations
2) Sector - wise impact being different and Art and cultural NonProfits being worst affected.
3) Rising influence of crowdfunding as a funding source for SMEs.

Step - 3: Analyzing alternate funding source - crowdfunding






main_cat
art
music
technology
publishing
food
fashion
games
design
comics
photography
crafts
theater
dance
journalism
film & video
main_category
Music
Publishing
Games
Art
Design
Technology
Food
Fashion
Comics
Theater
Photography
Crafts
Journalism
Dance
Film & Video

Most and Least Successful

Successful -

indie rock, country & folk, rock, shorts, accessories, fiction, comic books, documentary, children's books, nonfiction, with over 90% in that category being successful.

Failed -

mobile games, video, photo, couture, web, translations, r&b, software, food trucks, action, with around 70 - 80% failure.

2009-2018: Campaign performances

	Failed	Successful
Top campaign categories (by percentage of total)	Video, Food Trucks, Apps, Candles, Web, Farmers Markets, Restaurants, Hip-hop, Software, Mobile Games	Chiptune, Residencies, Anthologies, Dance, Indie Rock, Letter Press, Country and Folk, Classical Music, Theatre and Performances

Key Observation

Despite High support for Art and Culture sector on Crowdfunding platform like KickStarter, still due to lack of usage by NonProfits, it remained as largely an unexplored fund source.

Also, its popularity remains very high in US compared to other countries and even during covid, the proportion of projects by countries didn't seem to change much

So, due to similarities in countries and patterns of categories, we conclude that donor patterns remains the almost the same.

The Data source:
1) Kaggle
A. https://www.kaggle.com/kemical/kickstarter-projects
B. https://www.kaggle.com/sripaadsrinivasan/kickstarter-campaigns-dataset

2) Data from Scrapper, available here:
https://webrobots.io/kickstarter-datasets/

Practical:

1. ML to observe performance of feature selection.
2. ML to Predict the outcome of crowdfunding campaign
3. Analyzing Cycle Dataset to find insights for marketing campaigns (period: December to May)

Supervised Classification

Algorithm	Internal Data	External Data
SelectKBest	0.53	0.60
Mutual_info_classif	0.75	0.75
Forward Feature selection	0.75	0.77
SelectFromModel with Random forest estimator	0.77	0.86
Nystroem	0.79	0.65
RBF-Sampler	0.65	0.65

Un - Supervised Classification

Algorithm	Internal Data	External Data
Correlation	0.60	0.58
PCA	0.74	0.91
Mean Shift	0.74	0.88
Dispersion rate	0.69	0.70

Numeric variables

Algorithm	Accuracy
RandomForest	0.87
Complement Naive Bayes	0.70
Multinomial Naive Bayes	0.72
Stochastic Gradient descent classifier	0.76
Multi-layer Perceptron classifier	0.88

Name

Algorithm	Accuracy
RandomForest	0.68
Complement Naive Bayes	0.60
Multinomial Naive Bayes	0.61
Stochastic Gradient descent classifier	0.69
Multi-layer Perceptron classifier	0.69
BERT	0.61

Blurb

Algorithm	Accuracy
RandomForest	0.60
Complement Naive Bayes	0.61
Multinomial Naive Bayes	0.61
Stochastic Gradient descent classifier	0.62
Multi-layer Perceptron classifier	0.62
BERT	0.61

Combined

Algorithm	Accuracy
RandomForest	0.623
Complement Naive Bayes	0.724
Multinomial Naive Bayes	0.630
Stochastic Gradient descent classifier	0.773

Cycle dataset findings:

1. Annual members didn't ride docked bikes post January

   Docked bike usage grew among casual members after February.

2. Classic bikes were most used, (exception - in December casual users rode more electric bike)
3. Least travel occurred during February

   Travels decreases Dec through Feb (lowest) and increases significantly Mar through May.

   Spring - Start Summer: within Lake Shore Dr & Monroe St, followed by Streeter Dr & Grand Ave. Lake Shore Dr & Monroe St and Streeter Dr & Grand Ave have high degree centralities in spring and May.

   Winter: Ellis Ave & 60th St to Ellis Ave & 55th St and back and within Dearborn St & Erie St. Clark St & Elm St has high degree centralities.

   Lake Shore Dr & Monroe St. are important connectors by betweenness scores.

Month	Stations A to B	Frequency (for all types)
December	Within Lake Shore Dr & Monroe St	188
	Within Dearborn St & Erie St	140
January	Ellis Ave & 60th St to Ellis Ave & 55th St	182
	Within Dearborn St & Erie St	148
February	Ellis Ave & 60th St to Ellis Ave & 55th St and back	86
		82
March	Within Lake Shore Dr & Monroe St	637
	Within Streeter Dr & Grand Ave	456
April	Within Lake Shore Dr & Monroe St	963
	Within Streeter Dr & Grand Ave	768
May	Within Streeter Dr & Grand Ave	1786
	Within lake Shore Dr & Monroe St	1264

Total

Month	Stations A to B	Frequency (for all types)
December	Within Lake Shore Dr & Monroe St	174
	Within Millenium Park	68
January	Within Lake Shore Dr & Monroe St	56
	Within Michigan Ave & 18th St	43
February	Within Lake Shore Dr & Monroe St	28
	Within Millenium Park	27
March	Within Lake Shore Dr & Monroe St	587
	Within Streeter Dr & Grand Ave	395
April	Within Lake Shore Dr & Monroe St	874
	Within Streeter Dr & Grand Ave	657
May	Within Streeter Dr & Grand Ave	1607
	Within lake Shore Dr & Monroe St	1137

Table 12: Depicting trends for casual users

Casual