US Banks Hiring Trend Analysis

Summary	Conduct a study on the job openings in the top US Banks and analyze trends in the industry particularly in the area of Fintech
URL	https://github.com/pratikshsawant5293/ADS_Spring_2019/tree/master/Assignment_1
Category	Web
Environment	NA
Status	Version 1
Feedback Link	https://github.com/pratikshsawant5293/ADS_Spring_2019/tree/master/Assignment_1
Team No	7
Team Members	Maneendar Sorupaka (sorupaka.m@husky.neu.edu) Pratiksha Sawant (sawant.prat@husky.neu.edu) Rajsharavan Senthilvelan (senthilvelan.r@husky.neu.edu) Reema Mehta (mehta.r@husky.neu.edu)

Introduction

The merger of technology with finance/banking domain has introduced a new phenomenon of Financial Technology a.k.a. FinTech and its implementation is evidently visible in several financial institutions in the US. These institutions are coming out of mainstreams to implement technologies like data science, AI, Cloud Computing, Machine Learning and are capitalizing heavily to upgrade their technological platforms. Some of these enactments shared by fintech magazine highlight on the key areas financial institutions are developing.

Fintech companies are providing multiple career opportunities which makes it interesting to understand the hiring trends of top Banks in the United States. Our case study analyzes these trends in the Fintech industry to understand the mandate of technology.

Subject of Analysis : State Street and BB&T Corp Bank

Identifying Fintech Keywords

To begin with we first identify the prominent keywords typically used in Fintech. We refer the reports summarized by World Economic Forum (WEF) which discuss the state of the industry and their upcoming trends.

- [Beyond Fintech: A Pragmatic Assessment Of Disruptive Potential In Financial Services](#)
- [The Future of Financial Services](#)
- [The future of financial infrastructure](#)
- [A Blueprint for Digital Identity](#)

We extract the top 100 keywords and build three lists using the following approaches:

1. Wordcount

Wordcount counts all words in the pdfs and ranks them by frequency of appearance. The word cloud below highlights keywords identified by this approach.

2. TF/IDF (Term Frequency/ Inverse Document Frequency)

In information retrieval, tf–idf or TFIDF, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. It is often used as a weighting factor in searches of information retrieval, text mining, and user modeling. The tf–idf value increases proportionally to the number of times a word appears in the document and is offset by the number of documents in the corpus that contain the word, which helps to adjust for the fact that some words appear more frequently in general. Tf–idf is one of the most popular term-weighting schemes today; 83% of text-based recommender systems in digital libraries use tf–idf.

The word cloud below highlights keywords identified by TF/IDF.

3. TextRank

TextRank which is adapted from PageRank algorithm provides a score for each sentence in a text, out of which we take the top-n sentences and sort them as they appear in the text to build an automatic summary. We extract words instead of sentences and follow the same algorithm.The word cloud below highlights keywords identified by TextRank

After closely looking at all the word clouds we could determine that TF/IDF gives an accurate result of returning some relative fintech words and hence we decide to use word list derived from TF/IDF in next step of analysis.

Web Scraping

We built a scraper using selenium and BeautifulSoup for scraping career websites of BB&T Corp and State Street bank websites and parse each job posting for keyword extraction. These keywords are compared with the list generated from TF/IDF algorithm to find out the frequency of the fintech words among all the job postings respectively.
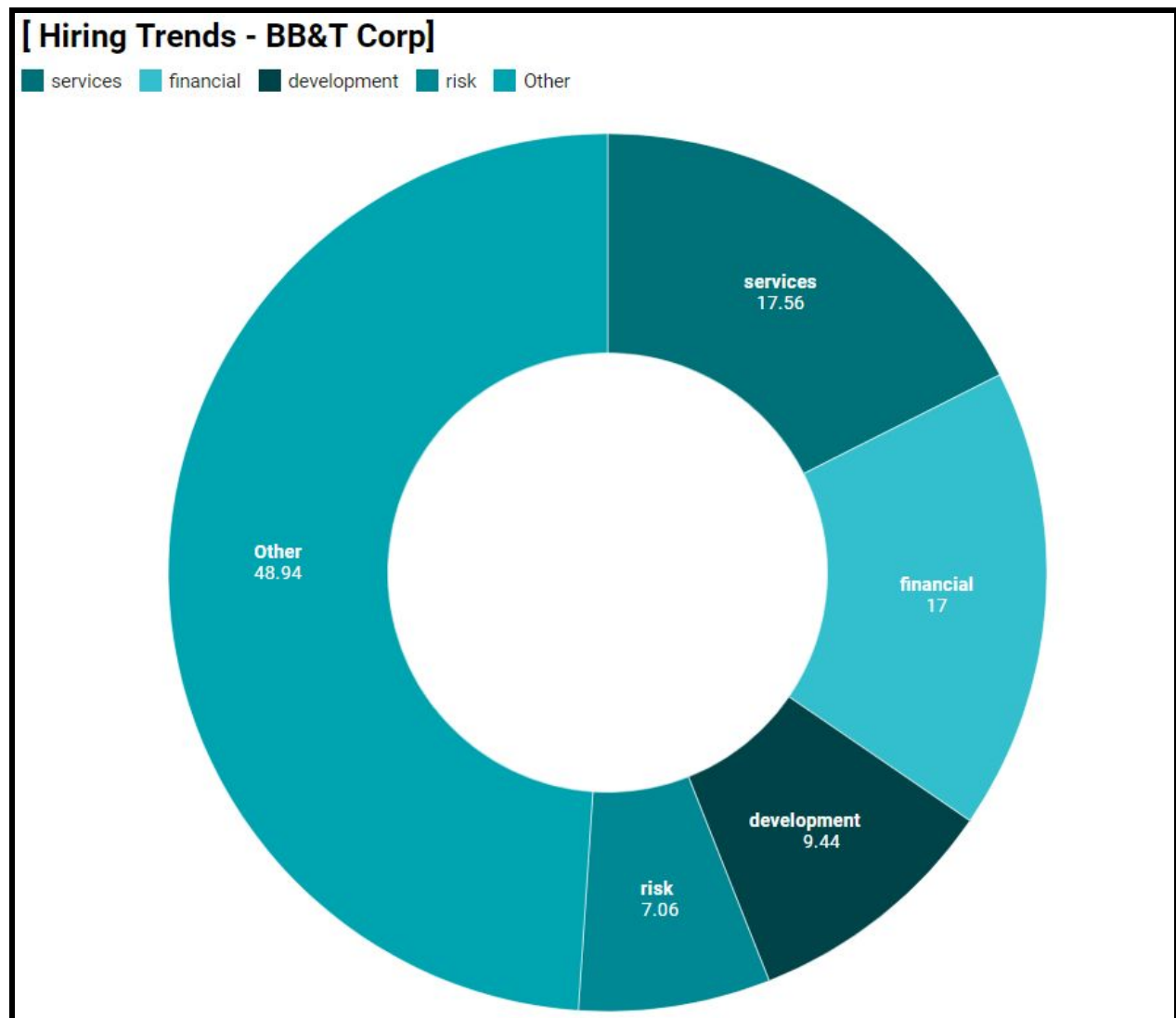
Key feature: We worked on dynamically generated web pages which are JavaScript rendered pages, content that is not present in the source code of the page. In order to retrieve URLs for each job posting, we deduced a logic on how the URL is built and identified a common pattern among them in turn scraping all the data from each of the job posting.

```python
for link in test_links:
    designation = link.find('div',attrs={'class':'gwt-Label WOTO WISO'})
    address = link.find('span', attrs={'class':'gwt-InlineLabel WM-F WLYF'})
#    print('designation is: ', designation.get_text())
#    print('address is: ', address.get_text())
    address_label = address.get_text()
    title = designation.get_text()
    address_label_contents = address_label.split('|')
    print(address_label_contents)
    print('length of list: ', len(address_label_contents))
    location_more = address_label_contents[0].strip()
    if location_more.count(',')>1:
        location_trim = location_more.rsplit(',',1)
        location = location_trim[0]
    else:
        location = location_more
    location = re.sub('[^a-zA-Z0-9-]+','-',location)
    title = re.sub('[^a-zA-Z0-9-]','-',title)
    job_id = address_label_contents[1].strip()
# #    Location.replace(',','-')
# #    print(Location, job_id)
    if len(address_label_contents)==2:
        url = base_url+title+'_'+address_label_contents[0].strip()
    else:
        url = base_url+location+'/'+title+'_'+job_id
    print(url)
```
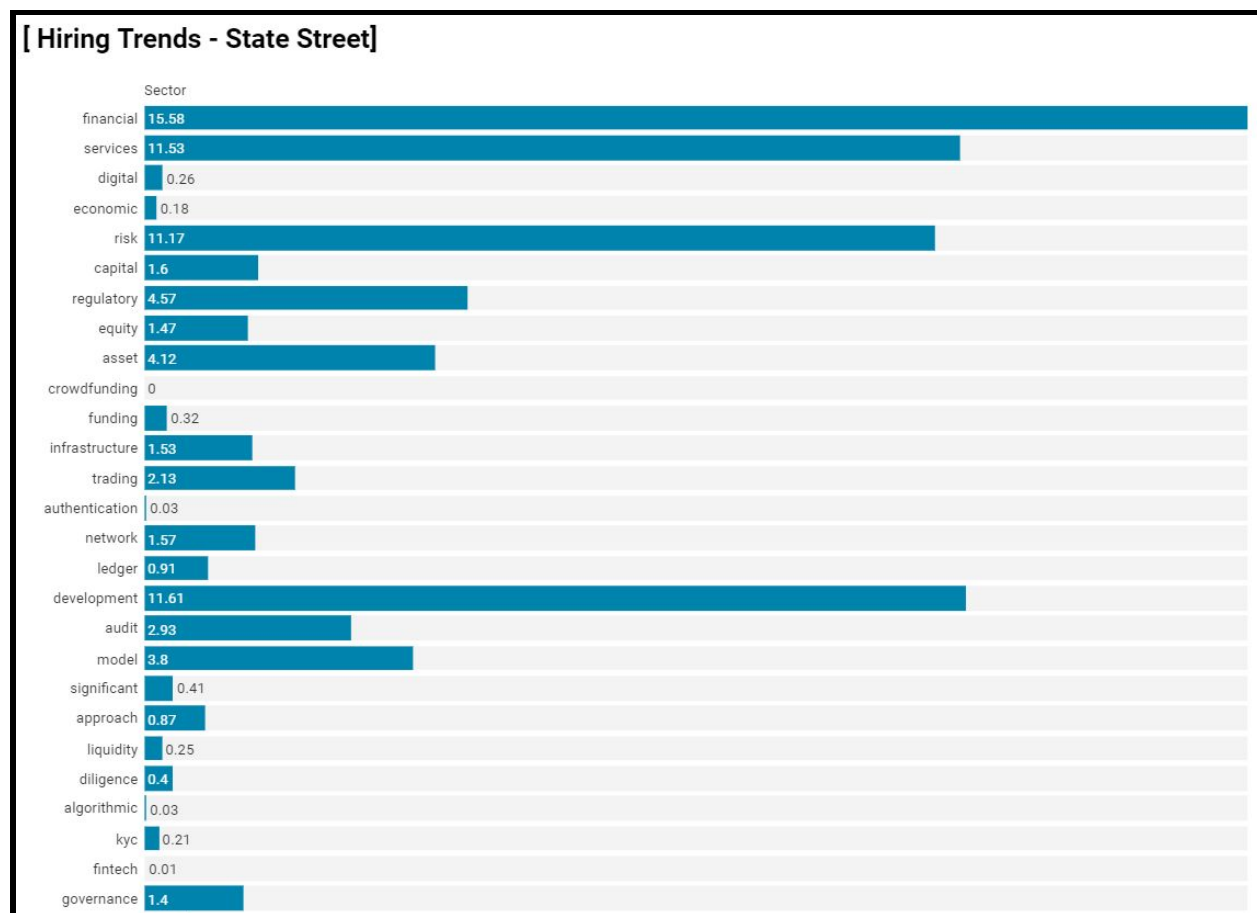
Analysis

BB&T Corp

From the data visualization represented below, BB&T Corp main focus is on services, development, risk and financial business sectors by hiring professionals who has expertise in these areas.

[Hiring Trends - BB&T Corp]
services financial development risk Other

services 17.56
financial 17
development 9.44
risk 7.06
Other 48.94

State Street

From the data visualization represented below State Street is chiefly concentrating on services, development, risk and financial business sectors by hiring professionals who has expertise in these areas. But it is also looking for people who are skilled in analytics, cloud, network and infrastructure which represents that it is looking forward to incorporate the upcoming technologies in its business.

[Hiring Trends - State Street]

Sector	Value
financial	15.58
services	11.53
digital	0.26
economic	0.18
risk	11.17
capital	1.6
regulatory	4.57
equity	1.47
asset	4.12
crowdfunding	0
funding	0.32
infrastructure	1.53
trading	2.13
authentication	0.03
network	1.57
ledger	0.91
development	11.61
audit	2.93
model	3.8
significant	0.41
approach	0.87
liquidity	0.25
diligence	0.4
algorithmic	0.03
kyc	0.21
fintech	0.01
governance	1.4

Conclusion

From the above observation we infer that banking industry is currently focused on hiring people skilled in the services, development, risk and financial business sectors. However, it is gradually bending towards implementing financial technology such as cloud, analytics, digitalization to automate their working platforms.

Citations

https://github.com/JRC1995/TextRank-Keyword-Extraction/blob/master/TextRank.ipynb

http://www.ultravioletanalytics.com/blog/tf-idf-basics-with-pandas-scikit-learn

https://medium.freecodecamp.org/how-to-process-textual-data-using-tf-idf-in-python-cd2bbc0a9
4a3

https://michaeljsanders.com/2017/05/12/scrapin-and-scrollin.html

https://relatedwords.org/relatedto/fintech

https://www.datawrapper.de/

https://worditout.com/word-cloud/create

Languages and Tools Used

Language	Python 3.7
Libraries/Tools	pdfminer, nltk, selenium, BeautifulSoup, Chromedriver
Visualization	Datawrapper, worditout