

Practicum Update for Attain,LLC

By Weihang Wen, Mengchen Xiao, Xuan Yang, Weichao Zhu, Yinian Lyu

**Master of Science in Business Analytics, Sep 2018,
George Washington University School of Business**

A Update submitted to

The Faculty of

The School of Business

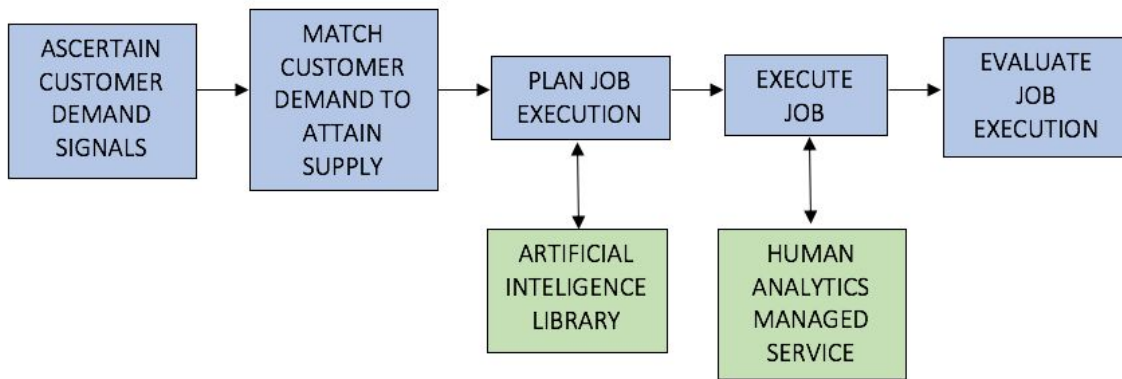
Of The George Washington University

In partial fulfillment of the requirements

For the degree of Master of Science

In Business Analytics

Sep 24, 2018



Flow of Attain Answers

Problem Description

The problem we are trying to solve is to help Attain. Inc, improve its digital consulting platform. Attain. Inc, a risk consulting company, now is planning to build a brand new consulting platform based on Attain Answers which is a cloud-based collaborative work environment. By leveraging the Attain Answers, we first can import and store the customers' profile, services that Attian might provide and project information and then build project plan, divide the plan into small steps and assign them to appropriate ones. Some jobs will be executed by Artificial Intelligence library and the place we kick in is to build Human Analytics Managed Services used to analyze public attitude and reflection towards certain topics which may lead to risks. Thus, to identify risk, what we do is to build a pipeline connecting a web scraping platform called Webhose and social network analysis platform called Maltego. After finished the pipeline, we will be working on Anti-semitism & Antifa topic and analyze how people who against Anti-semitism & Antifa connecting with each other or other organizations and what risk they might cause. The last part of the project is to evaluate the job execution. After the project finished, all these information will be stored in database and used to analyze customers' demand trend by using unsupervised learning and what we do is to improve Attain consulting services by using that trend.

Besides, to better and faster execute the plan, the company needs to choose the right members. This is another place we will be part of it. We will index everyone's skills on their resumes and find the one with skills suitable for different plans, by utilizing the AI library.

The another problem is that will risk factors e.g, schedule, budget, etc. influence the overall project risk and if they do, what is the relationship between them. The dataset Attain provides us with 8 projects and two aspects of risk in every project. One aspect focuses on difference in risk assessment between project manager risk and Attain assessment. Another one shows the way risk assessment project manager obtained by timing the rates of probability and impact. With the help of risk data, we are able to see the upcoming risks and prevent risks with high impact and high probability during planning job execution phase and job execution phase.

Questions we are going to answer

Anti-semitism & Antifa trends across the sites

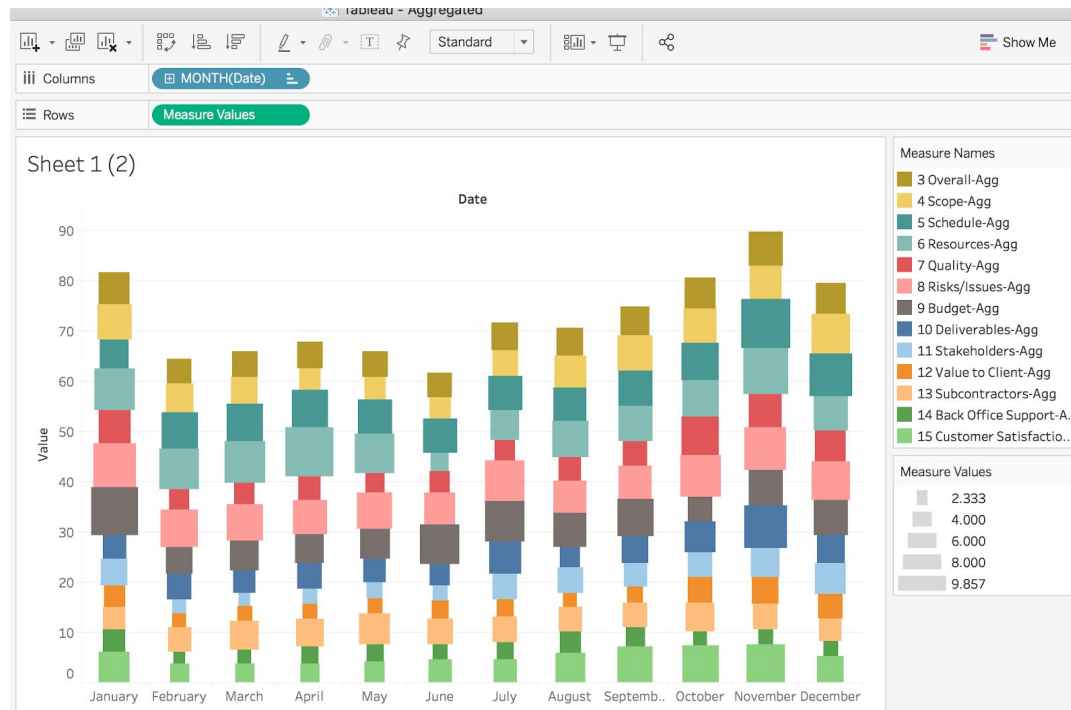
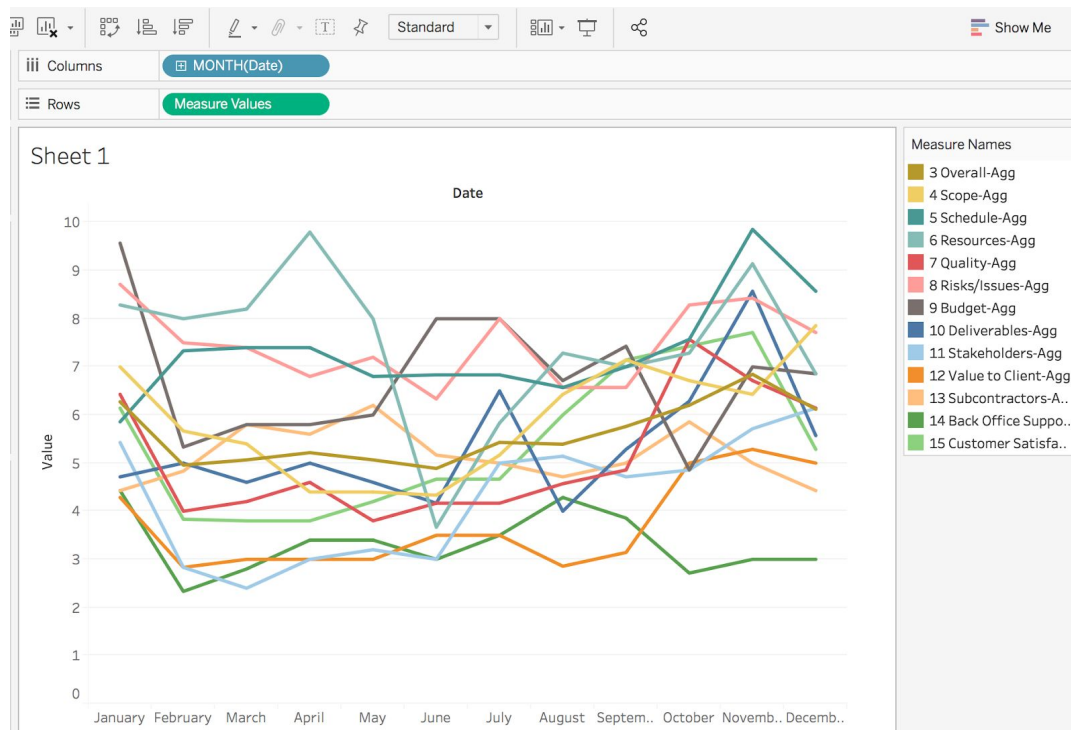
One topic we are working on is Anti-semitism & Antifa trends across the sites. Customer of Attain is specifically interested in understanding Anti-semitism & Antifa trends across the sites. Since both of these trends can cause violence, customers really want to understand how these trends are developing and whether they are trending toward violent expression. However, typically our customer would conduct research on extremism groups by reading news, blogs or articles from website to website, which might be time-consuming and might make some underlying connections to be overlooked. What Attain wants to do is to look at these groups from a more macro perspective and to provide insights that might be imperceptible when looking from only one aspect.

To be more specific, we want to figure out the leaders and members in these extremism groups, and the platform that they use to communicate and exchange information within the group. We would also look into major movements they have organized, focusing on their ideology, strategy, plans, and both physical and nonphysical assets. In addition, utilizing social network analysis, we would analyze the relationship between different entities involved to see whether they are concentrated in one area or connected by one specific person or organization. Besides, we are also interested in how dangerous these groups are. We would detect and then examine their threat language and also analyze their energy and physical assets to figure out their ability of turning into actual violence so that we could identify risks in advance.

Risk data

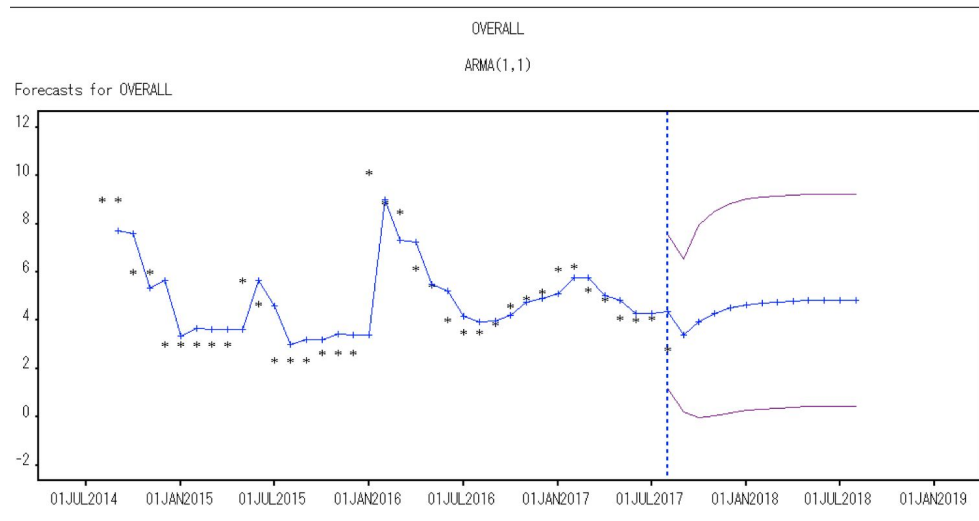
Another topic we are working on is the to understand the risk within Attain. We examined Attain risk data to understand if it is possible to predict recurring risks from the risk factors provided. The risk data we got has the probability, impact and criticality for 8 different projects overtime and each project has the same risk factors such as resources and quality. Here, the criticality

equals probability times the impact. We want to see if there is a pattern in the risk data and those risks can be predicted by a model. Therefore, the company can know on which month the risk for the project will increase and be well prepared for those risks.



When we made different trends for each project based on the criticality overtime by using the tableau, we found that there might be a seasonality or time series in the risk

data. Since we are trying to predict those risks, we made time series models for the overall risk and risks for different projects.



Forecast Data Set

OVERALL
ARMA(1,1)

	ACTUAL	PREDICT	U95	L95	ERROR	NERROR
01APR2017	4.8600	5.0269	8.2123	1.8414	-0.1669	-0.1027
01MAY2017	4.1100	4.8360	8.0215	1.6506	-0.7260	-0.4467
01JUN2017	4.0000	4.2988	7.4842	1.1133	-0.2988	-0.1838
01JUL2017	4.1100	4.2892	7.4746	1.1038	-0.1792	-0.1103
01AUG2017	2.8000	4.3731	7.5585	1.1877	-1.5731	-0.9679
01SEP2017	.	3.3780	6.5634	0.1926	.	.
01OCT2017	.	3.9473	7.9308	-0.0363	.	.
01NOV2017	.	4.2974	8.5439	0.0509	.	.
01DEC2017	.	4.5127	8.8545	0.1709	.	.
01JAN2018	.	4.6451	9.0225	0.2678	.	.
01FEB2018	.	4.7266	9.1172	0.3359	.	.
01MAR2018	.	4.7767	9.1724	0.3809	.	.
01APR2018	.	4.8075	9.2051	0.4098	.	.
01MAY2018	.	4.8264	9.2247	0.4281	.	.
01JUN2018	.	4.8381	9.2367	0.4394	.	.
01JUL2018	.	4.8452	9.2439	0.4465	.	.
01AUG2018	.	4.8496	9.2484	0.4509	.	.

Statistics of Fit

OVERALL
ARMA(1,1)

Statistic of Fit	Value
Mean Square Error	2.24504
Root Mean Square Error	1.49835
Mean Absolute Percent Error	22.72765
Mean Absolute Error	0.96447
R-Square	0.440

Here are the screenshots from the SAS. We did the time series for the overall risks. The dots behind vertical line are those predictions. The R-Square for the overall risk is 44%, which means 44% of the data is predicted by the model.

Some Correlation Test between Risk index input and Total Risk Rating

For the risk data, we also noticed the difference between the risk index input by the Product Managers and Peer-Reviewed input as Planned index and Actual Index. Holding the assumption that there could be some relationships between the difference of those two index and the total risk rating for each project. We conduct some correlation test for the two data.

```
> # shapiro-wilk normality test for risk difference
> shapiro.test(risk$total.risk.difference)
```

```
Shapiro-Wilk normality test
```

```
data: risk$total.risk.difference
W = 0.78405, p-value = 2.059e-06
```

```
> # Shapiro-Wilk normality test for risk rating
> shapiro.test(risk$Total.Risk.Rating)
```

```
Shapiro-Wilk normality test
```

```
data: risk$Total.Risk.Rating
W = 0.92191, p-value = 0.006978
```

Here are the normality test for two data input. The small p-value here demonstrated that the total risk difference and total risk rating do not conform normal distribution.

```
> cor1 <- cor.test(risk$total.risk.difference, risk$Total.Risk.Rating,
+                 method = "pearson")
> cor1
```

```
Pearson's product-moment correlation
```

```
data: risk$total.risk.difference and risk$Total.Risk.Rating
t = 0.37017, df = 40, p-value = 0.7132
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2499408  0.3560378
sample estimates:
      cor
0.05842917
```

```
>
> cor1$p.value
[1] 0.713208
> cor1$estimate
      cor
0.05842917
```

The Correlation test we did here displayed that the two input are not strongly correlated as their correlation based on the Pearson test here is only 0.058.

```

> #Project 5
> cor8 <- cor.test(risk5$total.risk.difference, risk5$Total.Risk.Rating,
+                 method = "pearson")
> cor8

Pearson's product-moment correlation

data: risk5$total.risk.difference and risk5$Total.Risk.Rating
t = -6.1153, df = 3, p-value = 0.008788
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9975899 -0.5286062
sample estimates:
      cor
-0.9621522

>
> cor8$p.value
[1] 0.008788468
> cor8$estimate
      cor
-0.9621522

```

Here are the correlation test we did for each specific project. For distinct project, the correlations would be totally different.

```

> cor11 <- cor.test(risk8$total.risk.difference, risk8$Total.Risk.Rating,
+                  method = "pearson")
> cor11

Pearson's product-moment correlation

data: risk8$total.risk.difference and risk8$Total.Risk.Rating
t = 14.947, df = 2, p-value = 0.004446
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7981068 0.9999116
sample estimates:
      cor
0.995554

>
> cor11$p.value
[1] 0.004445972
> cor11$estimate
      cor
0.995554

```

More correlation test would be included covering not only the total risk difference but also include index for each risk factors. We would like to see more correlations between those index such that we could better identify the pattern of the risk for the customers of Attain.

Resume Text Mining

Attain also wants to assign the projects or tasks to appropriate people more efficiently at the beginning of a new project. By utilizing Attain's AI library, we would conduct unsupervised learning on employees resumes to create index for skills so that each

resume could be tagged with one or more skills the employee has, which could help to find the right person with the skills in need.

Proposal Text Mining

Just like creating index for skills in resumes, Attain would also like to create index for topics in proposals. We would utilize Attain's AI library to conduct unsupervised learning on proposals and build a model that could extract the main topics in a specific proposal. By doing so, managers could save plenty of time and efforts of reading proposals that have dozens of pages.

Project Risk

None of us in the group have had any experience in business consulting area and we are not familiar with the digital database and platform wanted by the attain company and our mentors before the practicum.

Being exposed to the social network analysis is also part of the project risk since we should be using some techniques from social network analytics to better predict and identify the pattern of the data.

Besides, having to dealing with a software that is completely new to us is also one of the project risk. We will have to get familiar with Maltego in a relatively short period of time and then to generate valuable insights from it.

Description of the possible deliverables

For us, the practicum was initially separated into two areas. One is a pipeline of social network analysis platform. Another one is mainly the risk data descriptive and predictive analysis.

Expected Outcome

The platform based on Webhose API and Maltego would mainly work for the purpose to help Attain build Human Analytics Managed Services platform and thus finish the skeleton of entire digital consulting platform, and the result of risk data analysis could be used to provide some data-driven suggestions to the consultants in the attain company.