

XN PROJECT DELIVERABLE FEEDBACK

Integrated Experiential Learning



Northeastern University

ALY6080, SPRING 2020

XN PROJECT

XN PROJECT DELIVERABLE FEEDBACK

SUBMITTED BY: ADSAR SHIVANI, KC SANJOG, MISHRA ANUPREETA, THORAT
NIKHIL

EMAIL ID: adsar.s@husky.neu.edu, Kc.s@husky.neu.edu, mishra.anu@husky.neu.edu,
thorat.n@husky.neu.edu

NUID: 001399374, 001050752, 001765981, 001056767

SUBMITTED TO: PROF. MAX MASNICK

DATE: 06/17/2020

XN PROJECT DELIVERABLE FEEDBACK

Introduction

This assignment allowed us to explore the data we were given. After understanding the problem from a business perspective and understanding the pros and cons of our data, we now move towards data analysis. Our data had very little need for things like leveling data and assigning variables, however it required some data reshaping. After doing that we did some data analysis and tried to understand what our data was trying to tell us. We used R, Tableau, and Excel to do the same.

Analysis

To understand the data and dig a little deeper into our data, we take 3 steps:

Step 1 – Reshaping Data:

This answers the question of how we have done data preparation. This step was done completely in R. To be able to convert our data into a format that the machine understands, we reshape our skill data into two formats, namely, long, and wide (Reshaping “Nontraditional” Wide Data (“Multi-Choice” Data) – Max Masnick, n.d.). We use the following process:

1. We load the packages and read the second sheet of our excel as it has the data needed for skill mapping, as seen in Fig. 1.
2. Now, as we can see from Fig 1, our data is represented in the non-traditional wide format. To make sense of this, we convert data into a long and wide format, so we

```
library(tidyverse)
library(janitor)
library(readxl)
library(sjmisc)
library(wordcloud)
library(tm)

getwd()
[1] "C:/Users/mishr/Google Drive/NEU/Quarter 5/XN"
setwd("C:/Users/mishr/Google Drive/NEU/Quarter 5/XN")
getwd()
[1] "C:/Users/mishr/Google Drive/NEU/Quarter 5/XN"

# import courses and skills file
course_skills_raw<- read_xlsx("XN Project_Skills by Course_4.2020.xlsx",sheet = 2)
dim(course_skills_raw)
[1] 45 41
# names
#> ...5
#> ...6
#> ...7
#> ...8
#> ...9
... and 32 more problems
course_skills_raw
# A tibble: 45 x 41
  Course Code Title Description Skill Labels ...5 ...6 ...7
  <chr> <chr> <chr> <chr> <chr> <chr> <chr>
```

Fig 1: Loading data into R

XN PROJECT DELIVERABLE FEEDBACK

need to use a *pivot_longer* function in the *tidyverse* library. However, to do that we need a part of the title of the columns with skills to be the same, as *pivot_longer* needs a ‘starts with’ variable, as **this is easier than listing the names of each column and more robust to changes** in excel. From Fig 1, we see that all consecutive columns of *Skill Labels* have ‘... and a number’ as a column name.

- As per our observations from Fig 1, we got numbers that have “...” prefix, as those are the ones, we want to replace with *Skill Labels*, as we can see in Fig 2.

```
> colnames(course_skills_raw)
[1] "Course Code" "Title" "Description"
[4] "Skill Labels -->" "Skill Labels 1" "Skill Labels 2"
[7] "Skill Labels 3" "Skill Labels 4" "Skill Labels 5"
[10] "Skill Labels 6" "Skill Labels 7" "Skill Labels 8"
[13] "Skill Labels 9" "Skill Labels 10" "Skill Labels 11"
[16] "Skill Labels 12" "Skill Labels 13" "Skill Labels 14"
[19] "Skill Labels 15" "Skill Labels 16" "Skill Labels 17"
[22] "Skill Labels 18" "Skill Labels 19" "Skill Labels 20"
[25] "Skill Labels 21" "Skill Labels 22" "Skill Labels 23"
[28] "Skill Labels 24" "Skill Labels 25" "Skill Labels 26"
[31] "Skill Labels 27" "Skill Labels 28" "Skill Labels 29"
[34] "Skill Labels 30" "Skill Labels 31" "Skill Labels 32"
[37] "Skill Labels 33" "Skill Labels 34" "Skill Labels 35"
[40] "Skill Labels 36" "Skill Labels 37"
```

Fig 3: Testing our function

```
> course_skills_long <- course_skills_raw %>% pivot_longer(starts_with("Skill Labels"), values_drop_na = TRUE)
> course_skills_long
# A tibble: 826 x 5
  Course Code Title Description name value
  <chr> <chr> <chr> <chr> <chr>
1 LDR 6100 Developing ~ Begins with the premise~ Skill L~ Team Lea~
2 LDR 6100 Developing ~ Begins with the premise~ Skill L~ Communic~
3 LDR 6100 Developing ~ Begins with the premise~ Skill L~ Global L~
4 LDR 6100 Developing ~ Begins with the premise~ Skill L~ Emotiona~
5 LDR 6100 Developing ~ Begins with the premise~ Skill L~ Servant ~
6 LDR 6100 Developing ~ Begins with the premise~ Skill L~ Leadersh~
7 LDR 6100 Developing ~ Begins with the premise~ Skill L~ Journals
8 LDR 6100 Developing ~ Begins with the premise~ Skill L~ Transfor~
9 LDR 6100 Developing ~ Begins with the premise~ Skill L~ Presenta~
10 LDR 6100 Developing ~ Begins with the premise~ Skill L~ Leadersh~
```

Fig 4: Converting raw data to long

```
for (i in 1:length(course_skills_raw_colnames)) {
  if (str_contains(course_skills_raw_colnames[i], "...")) {
    getrownumber[i] <- i
  }
}
getrownumber
s <- na.omit(getrownumber)
s
temp <- array()
# Then we put Skill labels in our data header
for (i in 1:length(s)) {
  temp[i] <- paste("Skill Labels ", toString(i))
  names(course_skills_raw)[s[i]] <- temp[i]
}
colnames(course_skills_raw)
```

Fig 2: Function to Convert title into our format

- Next, we check if our function worked, as seen in Fig 3, it did.

- Thereafter, we reshape this to long and wide using *pivot* function as seen in Fig 4 and 5.

```
> course_skills_wide <- course_skills_long %>%
+ mutate(check = 1) %>% # Used for creating a binary 1/0 variable for e
+ pivot_wider(
+   id_cols = "Course Code",
+   names_from = value,
+   values_from = checked,
+   values_fill = list(checked = 0) # Fills in 0 for any value not "check
+   ed" above by the 'mutate()' function
+ )
> course_skills_wide
# A tibble: 45 x 408
  Course Code Team Leadershi~ Communications Global Leaders~
  <chr> <dbl> <dbl> <dbl>
1 LDR 6100 1 1 1
2 CMN 6000 0 1 0
```

Fig 5: Converting long to wide

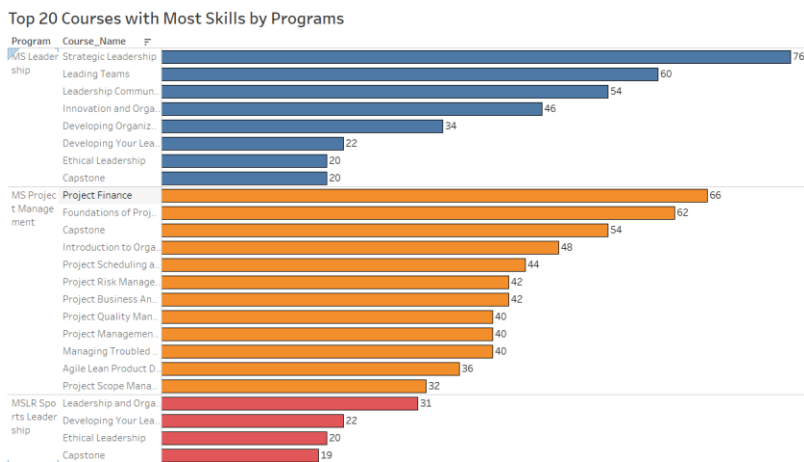
- These were then converted to CSV files that we could use in Tableau.

Step 2 – Exploratory Data Analysis:

This answers all the tasks that we have completed. This step was done in both Tableau and R.

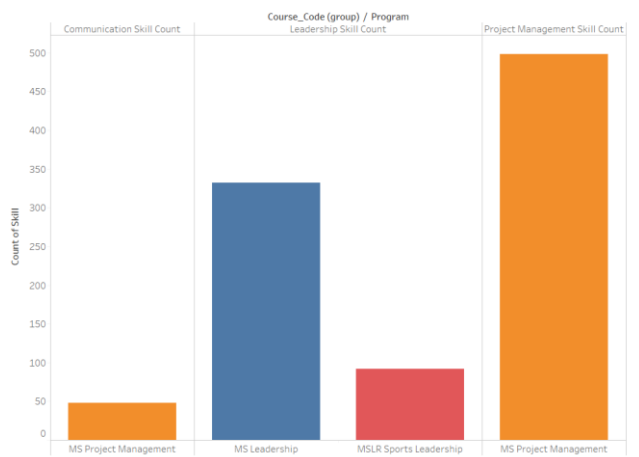
XN PROJECT DELIVERABLE FEEDBACK

Visualizations:



The bar graph shows the top 20 courses with most skills with respect to different programs. As we can see, Master's in Leadership has the top skill named, Strategic Leadership.

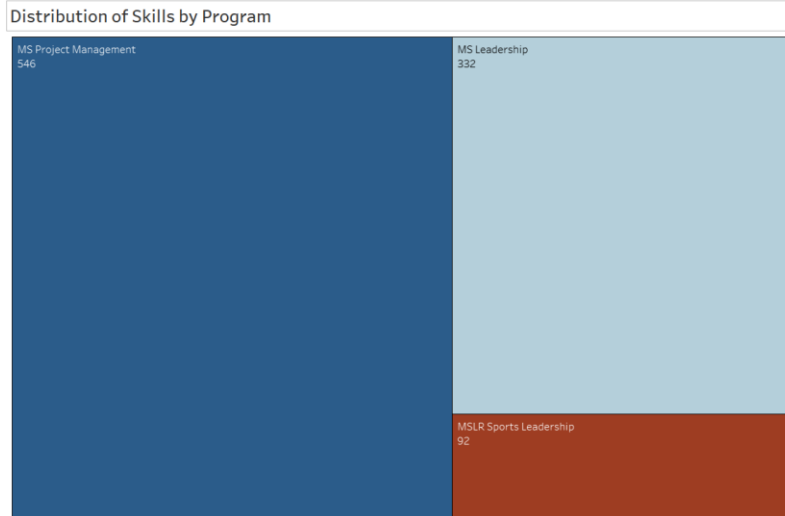
Fig.7: Top 20 Courses with most skills by programs



The bar graph shows the combination of skills within each of the three master's programs.

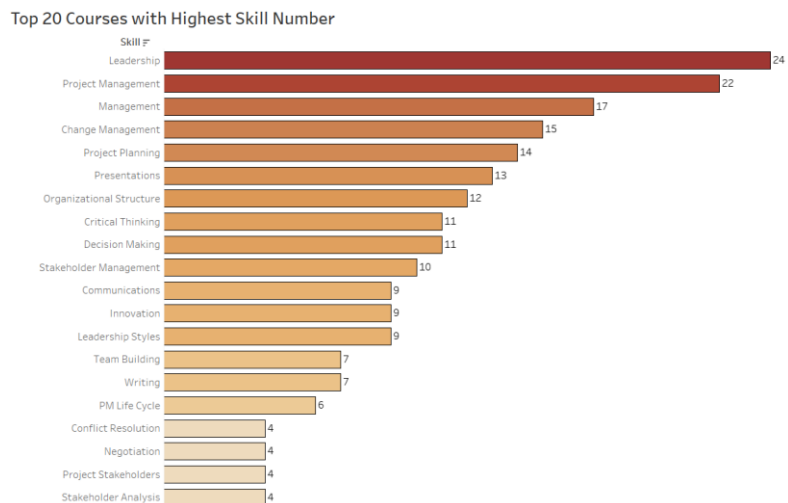
Fig.8: Combination of Skill type by program

XN PROJECT DELIVERABLE FEEDBACK



The tree map shows that 546 skills are delivered by MS Project Management program , 332 skills by MS Leadership and 92 in MSLDR Sports

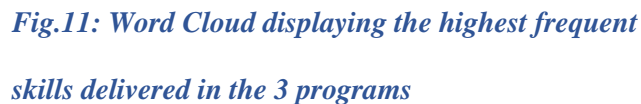
Fig.9: Distribution of skills by program



The bar graph shows, the top 20 courses with highest skill number. The Leadership course has the highest skill number.

Fig.10: Top 20 Courses with highest skill number

After this, we decided to perform text analysis to compare what we can see from the description of the skill to see if there was any relationship between them. As we can see from Fig 9 and 10, the most frequent word in both cases is Project, which shows that we want our courses to be practical. Then there is Leadership, which is the focus of our courses.



We have performed data analysis using the College of Professional Studies dataset, as given by our project sponsor. Initially, the reshaping of data was performed to prepare the data for analysis. Further, we researched some peer-reviewed articles to understand and gather views from various sources to improve business revenue. Finally, we have implemented visualizations on the data for demonstration to the client.

XN PROJECT DELIVERABLE FEEDBACK

Reference:

- [1] Reshaping “nontraditional” wide data (“Multi-choice” data) – Max Masnick. (n.d). Retrieved April 24, 2020, from <https://maxmasnick.com/kb/rlang/reshape-multi-choice/>
- [2] Börner, K., Scrivner, O., Gallant, M., Ma, S., Liu, X., Chewning, K., Wu, L., & Evans, J. A. (2018). Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. *Proceedings of the National Academy of Sciences*, 115(50), 12630–12637. <https://doi.org/10.1073/pnas.1804247115>