

Tarefa 4

Análise PCA

Nilson Furquim da Silva, 8937250

Pré-processamento

Primeiro os dados em .xls foram exportados para .csv. Depois todas as tuplas com dados faltando foram retirados do conjunto de dados. Então os valores de “TotalScore” foram modificados partir da 200ª posição de “-” para um valor dentro da escala de 0 a 100. Os valores foram alterados levando em consideração a média do intervalo e uma interpolação entre o “TotalScore” da 200ª universidade e zero.

Gráficos gerados

Ambos os gráficos gerados foram utilizando a forma de representação *star plot*. O primeiro mostrando cada universidade, com os valores mais claros e os gráficos maiores dispostos em ordem decrescente de classificação. O segundo mostra um gráfico cartesiano com cada gráfico em ordem decrescente de classificação e com altura baseada no “TotalScore”.

Análise

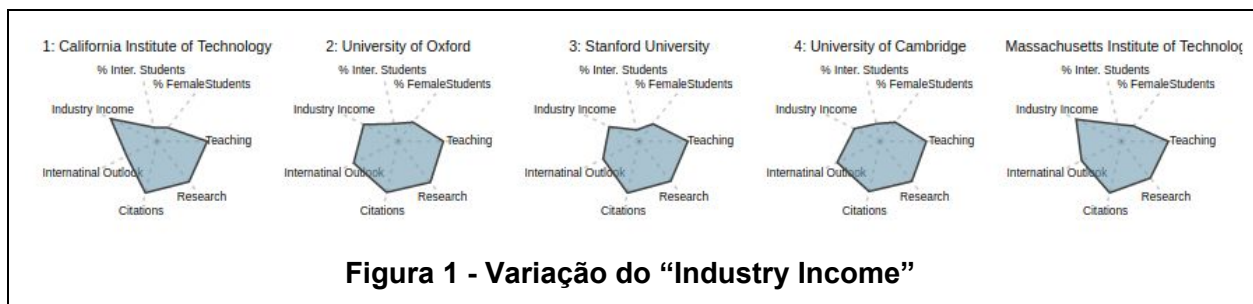
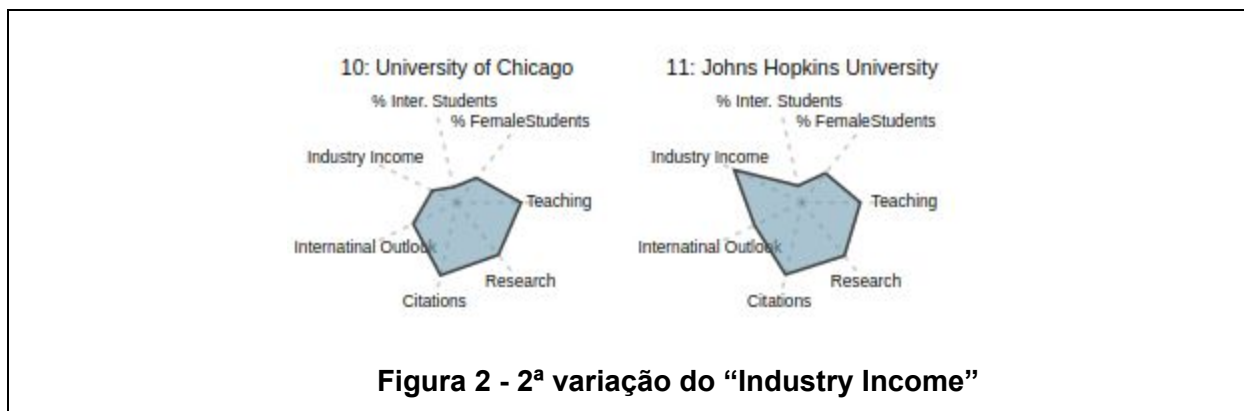


Figura 1 - Variação do “Industry Income”

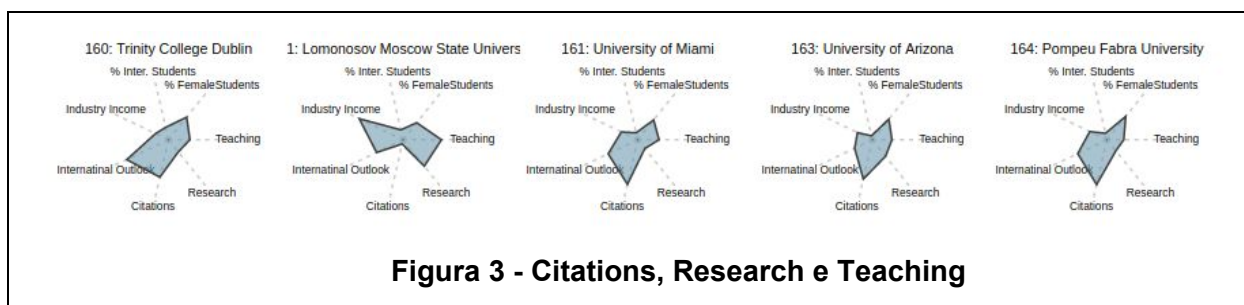
Observando a Figura 1, é possível observar a maioria dos atributos têm valores muito parecidos, tirando o “Industry Income”, que é mais alto na primeira e quinta posição levando a hipótese de que esse atributo influencia menos na classificação do que outros atributos.

A Figura 2 mostra mais um exemplo da variação do atributo “Industry Income” sem afetar a classificação da universidade. Nela temos na décima posição um valor muito mais baixo que a décima primeira posição. Reforçando que esse atributo influencia pouco a classificação da universidade no ranking.

Na Figura 3 há uma variabilidade maior dos atributos. Especificamente os atributos “Research”, “Citations” e “Teaching” variam mais entre as posições 161 e 162. Na posição 161 há um baixíssimo valor para “Citations” e um valor alto para “Teaching” e “Research”. Na



posição 162, um valor alto para citations baixo para “Research” e intermediário para “Teaching”. Levando a hipótese de que os 3 valores tem influência parecida na classificação.



Também é possível observar que classificações mais baixas, possuem “Citations”, “Research” e “Teaching” bem inferiores se comparados com as posições mais altas, mesmo com os outros atributos com valores altos. Além disso, é possível ver o padrão se repetindo de o “Industry Income” baixos com classificações melhores que “Industry Income” baixos.

Conclusões

Levando em conta as análises e o resto do gráfico é possível concluir que o atributo “Industry Income” tem pouca influência na classificação das universidades enquanto que os atributos “Citations”, “Research” e “Teaching” influenciam muito.