Term Project Report

COMMUNICATING THE RESULTS OF DATA ANALYTICS

CS587 – Spring 2017

Nitika Chaudhary (UBIT: nitikach, Person No.: 5020-9474)
Raj Jaysukh Patel (UBIT: rajjaysu, Person No.: 5020-8278)

Dataset Details

The data set is World University Rankings dataset from Kaggle about which universities in the world are the best? This dataset contains three global university rankings from very different places:

Times Higher Education World University Ranking (timesData.csv)

Academic Ranking of World Universities (shanghaiData.csv)

Center for World University Rankings (cwurData.csv)

Story of the Data

Following are the details about various sheets in our tableau workbook which tells you the story of the World University Rankings

Chart Name: ScoreVSQuality

X-axis: Longitude

Y-axis: Latitude

Data Source: cwurData.csv

Graph: Center for World University Rankings Score and Qualities(Education and Faculty)

Interpretations:

1. USA has high Score and high Quality of Education and high Quality of Faculty, according to Center for World University Rankings as compared to other countries.
2. Canada and United Kingdom has same score but United Kingdom has higher Qualities than Canada.
3. Japan has higher score than Netherlands but Netherlands has higher quality than Japan.

Chart Name: FeatureRanks

X-axis: Citations, Influence, Publications, Patents

Y-axis: Country

Data Source: cwurData.csv

Graph: Score of each feature for different countries, sorted by total score from high to low

Interpretations:

1. USA and United Kingdom have the highest total score as well as higher feature scored as well which aligns with the ScoreVSQuality inference.
2. Some universities have comparatively low scores for their features but they excel in one of their feature thereby leveraging its overall score and ranking like Sweden, Switzerland etc. While there are universities with uniform scoring in all features like Canada, Australia.

Chart Name: WorldRanking

X-axis: University Name

Y-axis: World Rank

Data Source: shanghaiData.csv

Filters: Year, Country

Graph: Average of World Rank for Universities over all the years and comparison with National Rank

Interpretations:

1. More universities from USA are at the top which aligns with the readings made in previous charts.
2. As we move right in the graph the bars keep getting darker depicting lower National Ranks along with a decline in the world rank.
3. Over the years the universities are sliding in their ranks within a country but at world level country remains at the same position.

Chart Name: GlobalScore

X-axis: UniversityName

Y-axis: GlobalAverageScore, cwurScore, timesScore, shanghaiScore

Data Source: inner join(shanghaiData.csv,timesData.csv,cwurData.csv) on UniversityName

Filters: Country

Graph: Comparison between the university scores from different sources and ranking as per the global average score

Interpretations:

1. For top universities, scores from all the sources are matching with each other and the globalAverageScore goes in accordance with them.
2. There are universities which are ranked higher by one data source and comparatively lower by the other, for instance, for University of California, Berkley the times Score is quite higher making it rise in the globalAvergaeScore ranking.

Chart Name: Ratios

Data Source: timesData.csv

Filters: Year

Graph: Growth of Staff size versus number of students

Interpretations:

1. As the circle gets larger it even gets darker describing that the student staff ratio increases with increase in number of students hence we can infer that the rate of increase in staff size is much less than the rate of increase in number of students.

Data Product

Chart Name: UniversityClusters

X-axis: Awards

Y-axis: Highly Cited Researchers, Number of papers in Nature and Science and Publications in Science and Social Science

Algorithm: K means Clustering using R Script

Interpretations:

1. Cluster 1("Blue") has the universities which has higher awards for Number of papers in Nature and Science and Publications in Science and Social Science.
2. Cluster 2("Red") has the universities which has higher awards for Highly Cited Researchers. But has very low Number of papers in Nature and Science.
3. Cluster 3("Green") has average awards for Highly Cited Researchers and Number of papers in Nature and Science.
4. So, we can interpret that those universities which have got more awards for Highly Cited Researchers have less publications and vice versa. This indicates that those universities which

have higher number of Highly Cited Researchers will have less publications but of higher qualities which can get them awards.

RScript for the k means clustering is cluster.ipynb

Details about clusters:

```
Cluster means:
  data.hici  data.ns data.pub data.award
1 54.453480 52.55385 63.33223   54.75458
2  9.128803 10.10736 31.96927    1.93091
3 25.986145 23.30120 48.87425   12.18645


Within cluster sum of squares by cluster:
[1] 301914.4 511200.7 527516.0
```

We can observe that it is clear that when all three feature have high score that is if there are more Highly Cited Researchers, Number of papers in Nature and Science and Publications in Science and Social Science then award will be more and this is cluster 1 which has low sum of squares. This indicate that it very confident that the if all features are high then there will be more awards. But if there are less publications in any of the category than it is less confident that awards will be less as we observed that if the quality is high and quantity is low even then there are chances that they got more awards.

Chart Name: FutureTopUniversities

X-axis: World Rank

Y-axis: Year

Data Source: Times Data

Filter: Universities

Computed data: Top 10 universities for next 5 years with their positions.

Interpretations:

1. The prediction shows that it is highly likely that top universities will remain same for next 5 years.
2. Universities like University of Oxford**,** University of Cambridge, University of California, Berkely tends to go below their average Rank in next five years, While Massachusetts Institute of Technology and Princeton University tends to go above their average rank.
3. The model has determined that there is a 90% likelihood that the value of sales will be within the shaded area for the next 5 years. And the range of shaded area is more for universities which has more ups and downs before like Stanford University and University of Oxford.