Mustafa Abdullayev

Factory Performance Project

AY6010 – Probability Theory and Introductory Statistics

Term: Fall 2019

Instructor: Tom Breur

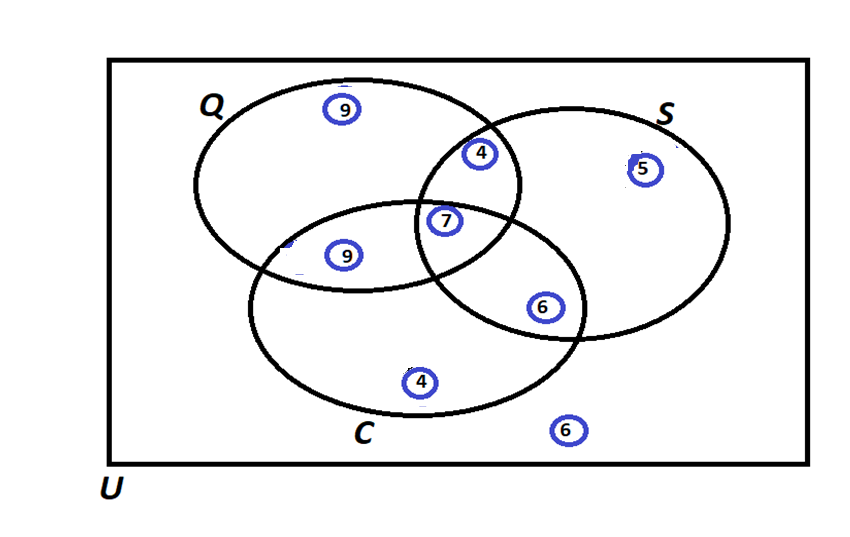
09/29/2019

Introduction

This is Microsoft Word Report accompanying Microsoft Excel Workbook. In my workbook, I analyzed dataset about untitled factory’s performance on recent projects. Data was provided by class instructor, Tom Breur. In total, it consisted of 50 observations about company’s recent projects. Each observation had 4 features – PIP number for uniquely identification, quality score of projects, process days and project cost. I used Empirical statistics and my main aim was to calculate probabilities of specific events with the help of given data. Data was cleaned, organized and ready for analysis. Also, I utilized powerful Excel built-in functions and graphs to dive deeper to observe hidden patterns and visually communicate my findings to audience.

Analysis

There are 3 main guidelines in this Project. First is that Project Quality is satisfactory is it is higher than 500. Then, it should take no longer than 13 days to be completed in order to classified as satisfactory. Finally Cost should be under 234.000 USD. Below, I put Venn diagram where I calculated all cases. Since there is a lot of conditional probabilities, visualizing makes calculating the given probabilities much easier. (Q: Quality, S:Speed,C:Cost).



In the exploratory analysis of data, I used box plots to see if there is any outliers in the given data and observed that there is no outliers.

a) Of those who satisfied cost, what percentage also satisfied Speed?

Notation : P(S | C) – Percentage of projects which has satisfactory length given that Cost is satisfactory?

Formula : P ( S ∩ C) / P (C)

Answer : 50 %

b) Of those who satisfied Quality, what percentage also satisfied Cost?

Notation : P ( C | Q ) – Percentage pf projects which has satisfactory cost given that Quality is satisfactory?

Formula : P ( C ∩ Q) / P (Q)

Answer : 55 %

c) Of those who satisfied Quality, what percentage also satisfied Speed but did not sat,sfy the Cost?

Notation: P( (S ∩ C' ) | Q ) - Percentage of projects with satisfactory speed and unsatisfactory Cost given that Quality is satisfactory

Formula : P ((S ∩ C') ∩ Q) / P(Q) = P ( S ∩ C' ∩ Q) / P ( Q )

Answer : 14 %

d) Of those who satisfied Cost, what percentage also satisfied Speed but did not satisfy the Quality?

Notation : P( (S ∩ Q') | C)

Formula : P ( ( S ∩ Q') ∩ C ) / P (C) = P (S ∩ Q' ∩ C ) / P(C)

Answer : 23 %

e) Of those who did not satisfy Speed, what percentage satisfied Quality and Speed?

Notation : P(( Q ∩ S ) | S')

Formula : P( (Q ∩ S ) ∩ S' ) / P(S') = P (Q ∩ S ∩ S' ) / P(S')

Answer : 0 % ( Since S ∩ S' = 0 )

f) What percentage satisfied exactly two of the three criteria?

Notation : P(Q ∩ S ∩ C') + P(Q ∩ S' ∩ C) + P(Q' ∩ S ∩ C)

Formula : same as Notation

Answer : 38 %

g) Of those who satisfied at least one of the three criteria, what percentage satisfied exactly one criterion?

Notation : ( P(Q ∩ S' ∩ C') + P(Q' ∩ S ∩ C') + P(Q' ∩ S' ∩ C) ) / (50 - P(Q' ∩ S' ∩ C'))

Formula : same as Notation

Answer: 41 %

h) Of those who did not satisfy Cost, what percentage satisfied the Speed criterion?

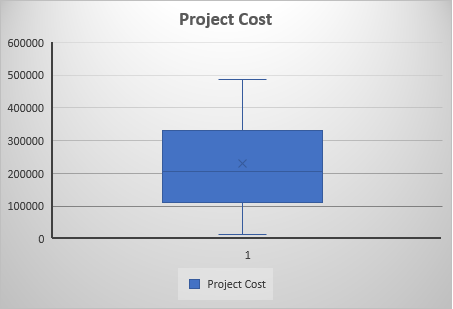
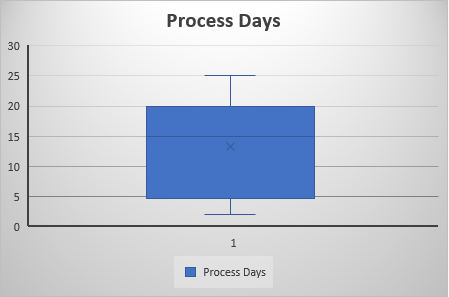
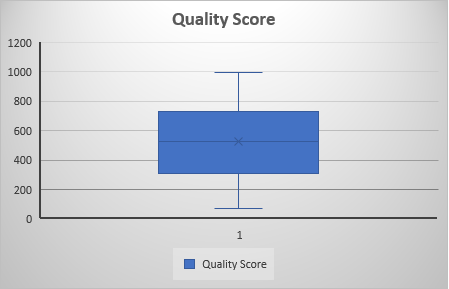
Notation : P(S | C')

Formula : P(S ∩ C') / P (C’)

Answer : 38 %

There are many interesting Facts about projects. First of all, more than hall of them failed to pass at least 2 criteria. Also, percentage of projects which passed one criterion was lower than 60 % for each criteria. That can signal two things. First it is possible that company has a major problem in their projects. On the other hand, it is possible that company’s criteria are very though.

On the other hand, I used Excel built-in charts to analyze data set. First of all, there is no outliers in any of three features. Also, means and medians are close to each other, tails are approximately same length. This shows that our features are approximately normally distributed.



Finally, it is reasonable to expect that there should be a correlation between Project Cost and Project Quality. But this assumption is falsified by Excel Scatter plot. Correlation is almost zero. Also, this is the same for Cost and Days. So, if quality is not correlated to Money spent on Project, then what variable can be explanatory ? In my opinion, Quality of team members is the first and foremost indicator of quality of Project. This can be analyzed using data about members. Data may contain years of experience, amount of domain knowledge, number of Project they have done and etc. Without further analysis we cannot say there is a correlation between quality of members and quality of Project. Even if there is a correlation, we should keep in my that, this does not imply causation – quality score is high due to members.

Conclusion

To conclude, I used Excel and Empirical Statistics to analyze Project Performance Data. Data had information about past 50 project of a untitled company. It had 4 features namely, PIP number, Quality Score, Process Days and Total Cost. I used given data to calculate probability of specific events. In order to facilitate our task, I utilized Venn diagrams. Key finding was that either company has very strict rules or company has problems in their project management. Moreover, intuitive correlation guess between Cost and Project Quality was not there. It was shown bu Scatter plot in Excel.