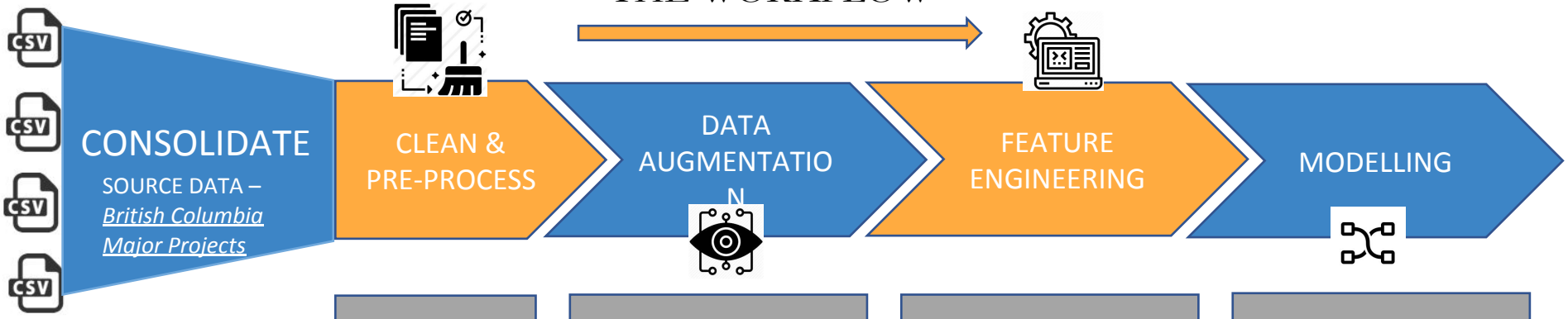
Ellisdon

Predicting Time and Cost Overruns

THE PROBLEM STATEMENT AND THE APPROACH

> Better predict the likelihood and impact associated with schedule delay and/or cost overrun risks on a large scale (~100M+) construction project.

THE WORKFLOW



- Replacing null values or dropping columns
- Creating Dummy Variables

- Added data to reflect unusual weather patterns, jobs generated
- Urban rural split for municipalities
- Public-Private and JV split for developers

- Identifying features with high correlation coeff and high statistical significance

- Generated 6 models including 2 each of linear regression, Decision Tree and Clustering Model

MODELS AND RESULTS

Linear Regression

Target Variables – Amount of delay (quarters), Amount of Budget Overrun

Model 1

	coef	std err	t	P>|t|	[0.0
CST_Mixed_Use	4.2709	0.904	4.727	0.000	2.4
Type_Owner_priv	2.9810	0.213	13.985	0.000	2.5
PUBLIC_FUNDING_IND	1.6891	0.274	6.163	0.000	1.1

Model 2

	coef	std err	t	P>|t|	[
Region_others	107.2531	33.481	3.203	0.001	4
CST_Roads_Highways	108.3863	35.568	3.047	0.002	3

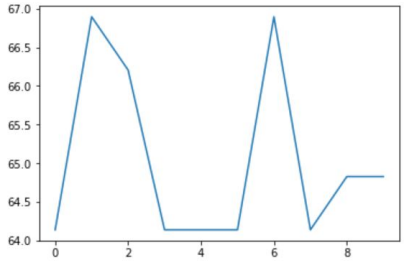
Results

Results are poor R sq 0.24 for model 1 predicting amount of delay, 0.03 for model 2 predicting budget overrun

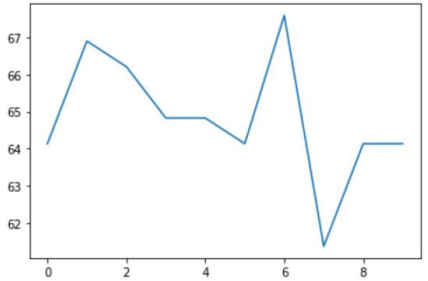
Decision Tree

Target Variables – Delay (yes / no), Budget overrun (Yes/no)

: [<matplotlib.lines.Line2D at 0x2ab8a609c88>]



[<matplotlib.lines.Line2D at 0x2ab8a665ac8>]



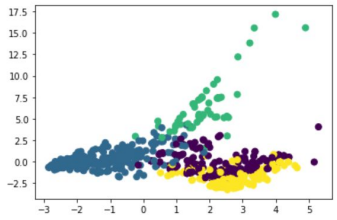
Results

Results are good, 66-67% accuracy for both target variables

Kmeans Clustering

Four clusters decided based on inertia metrics

: <matplotlib.collections.PathCollection at 0x2ab8bac9da6

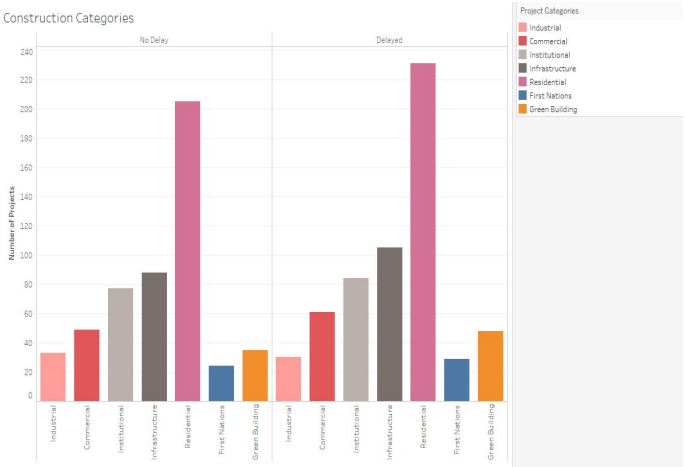


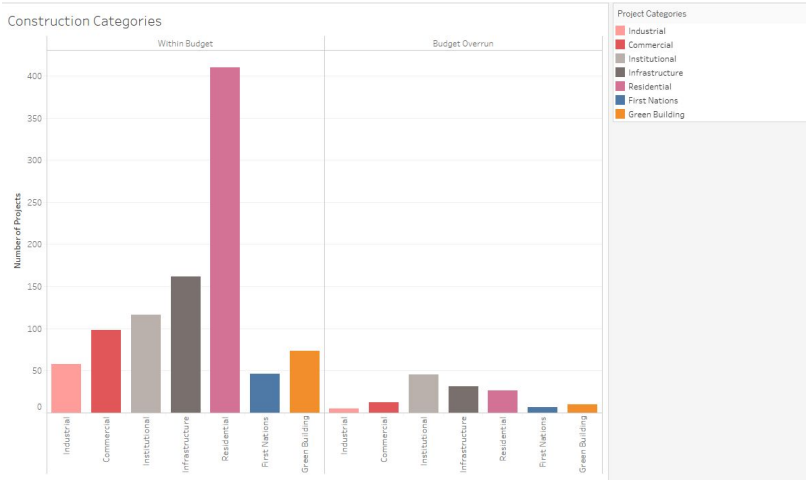
cluster	Budget_Overrun
0	0.161458
1	0.066298
2	0.086207
3	0.276471

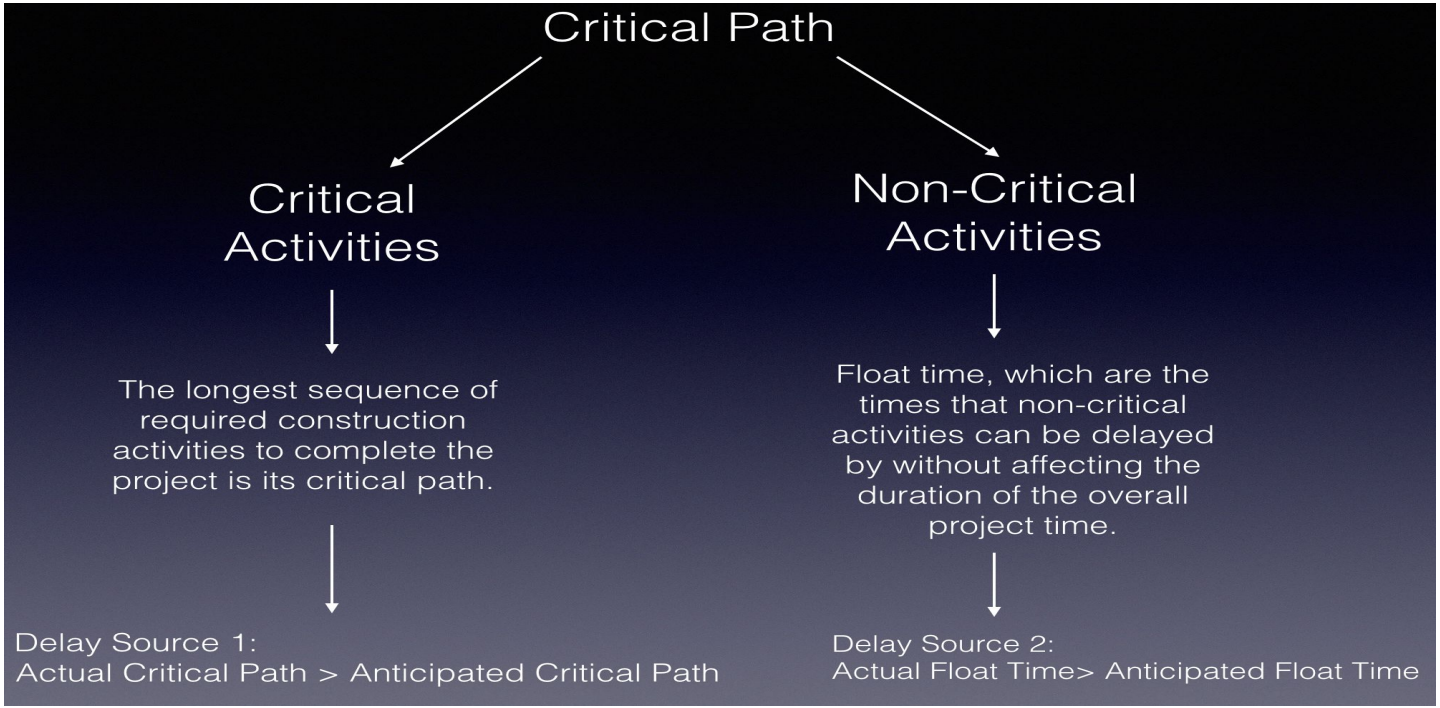
Results

Cluster 1 & 4 show high likelihood for budget overrun, 2&3 show low likelihood

Comparison Delays for Project Categories and Subcategories







Essentially, the ultimate goal is to optimize the time allocation between the critical path and float time.

Analyze critical path of past project, risk workshops in order to identify differences and similarities between projects that are delayed and projects that finished on time.

Steps Ellisdon can take for Better Data-Driven Decisions

The following is what Ellisdon has under their control and can use to avoid or significantly reduce overruns (Goal is to save/make money)

Addition of relevant features into predictive model

- Risk assessment value for different types of projects
- Contract type
- Number of third parties involved
- Predicted changes in inflation and cost of materials

Data Collection

- Ensure high quality and quantity of data collection
- Further large collection of surveys from construction managers, subcontractors & risk workshops

Conclusion

Not every delay is a bad delay

- Change in design by client would be compensated

There is a tendency to award project to lowest bidder with the shortest duration

- Some project types are doomed to be delayed & have cost overrun from the beginning

Expect overruns but focus on reducing the amount of overrun

- Using data-driven decisions