

Predicting Cancelled Tenders

Tuwaiq Four Week Bootcamp

Najwa Almulhem

What is happening?

- ✓ The Electronic Government Procurement System works to consolidate and facilitate the process of bidding and government procurement for all sectors in Saudi Arabia.
- On average **380** tenders are submitted per day.
- Each tender will need **2-24** man-hour for processing.
- **30%** of tenders are cancelled.

This costs is between 5k hours – 60K hours per month



Build a classification model to predict cancelled tenders



Better understand which factors are most important in cancelled tenders.

The Objective

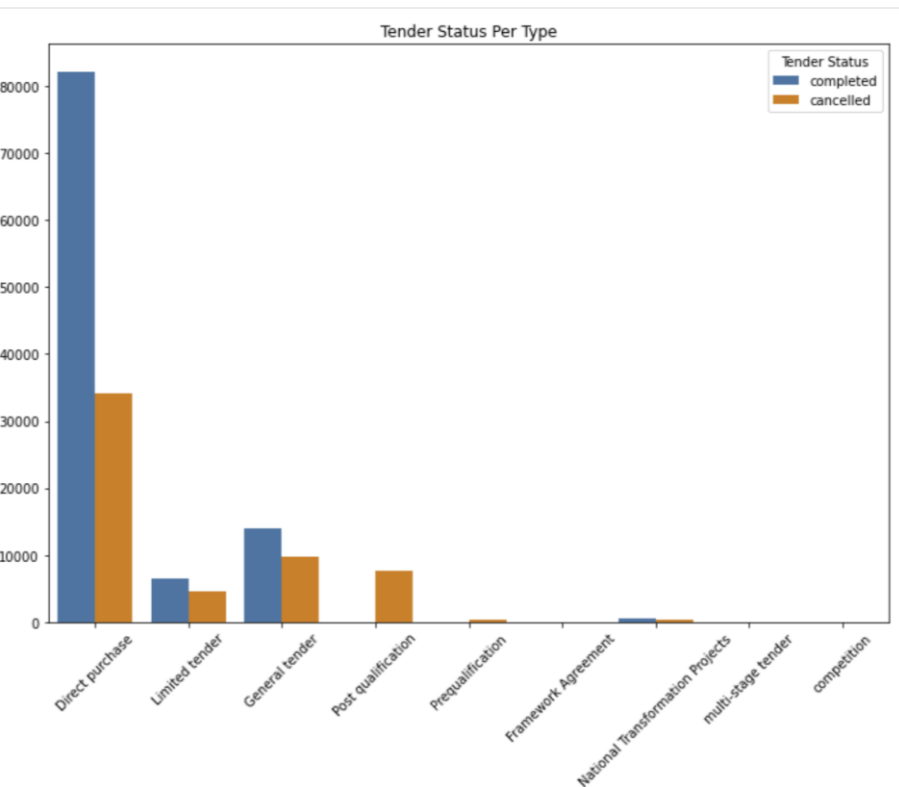
Data

Data was collected from Procurement System and masked for privacy protection

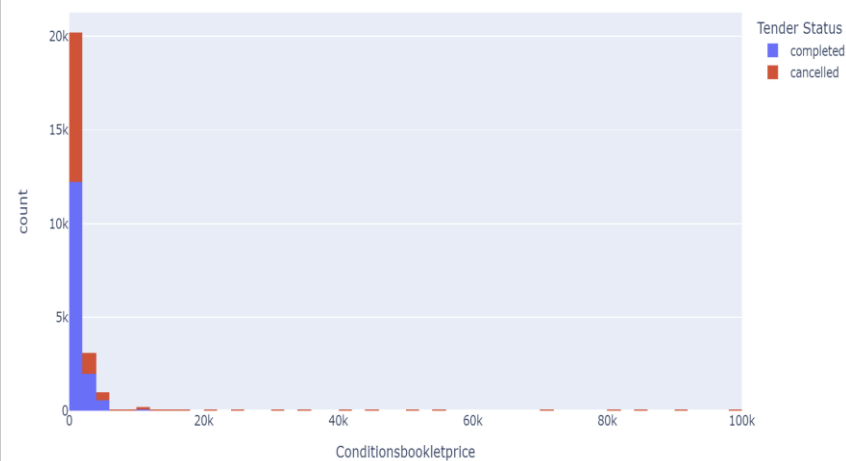
Format	Number of rows	Number of columns	Completed tenders	Cancelled Tenders
xlsx	160732	12	103629	57103

	Tender Type	Regions	Conditionsbookletprice	Tender Status	Invitation Type	Submissiondate	Lastofferpresentationdate	Activity Name	Estimatedvalue	Invit Count	categories_count	regions_count	latest_awarding_date
0	Direct purchase	city11	0.0	completed	private	2019-12-18 13:51:58.332	2019-12-23 00:00:00	Trade of tools, machines and devices	19444.444444	7	2	1	2020-02-13 13:18:27
1	Direct purchase	city11	0.0	completed	private	2019-12-18 13:51:58.332	2019-12-23 00:00:00	information technology	19444.444444	7	2	1	2020-02-13 13:18:27
2	Direct purchase	city11	0.0	completed	private	2019-12-22 12:53:07.624	2019-12-31 00:00:00	Insurance Activities	32777.777778	7	1	1	2020-02-13 13:12:54
3	Limited tender	city11	0.0	completed	private	2019-12-23 19:37:15.997	2019-12-26 00:00:00	Trade of tools, machines and devices	179444.444444	5	3	1	2020-02-20 17:47:36
4	Limited tender	city11	0.0	completed	private	2019-12-23 19:37:15.997	2019-12-26 00:00:00	Trading of utensils and household appliances	179444.444444	5	3	1	2020-02-20 17:47:36

EDA



Conditions Booklet Price with values 100-100K SAR



Word Cloud of all Tenders



Word Cloud of Cancelled Tenders



EDA

Based on EDA we can find :

1. Looking at tenders with Invitations higher than 50, we can see more tenders are cancelled.
2. Tenders with Conditions booklet price between 1-100 SAR , have almost equal probability based on the data for completion or cancellation.
3. Tenders with no Conditions booklet price have more probability of being completed based on the data.
4. General tenders have almost equal probability of being cancelled.

The Process

✓ Feature Engineering:

- Converting regions fields into columns since the region field can contain multiple regions or cities.
- Removing dates as it does not affect cancellation.
- Convert Invitation count , estimated value & condition booklet price into buckets or bins to increase accuracy.
- Do one-hot encoding for categorical variables.

✓ Modeling:

- Split Data into train ,test & validation.
- A model of the mean was set as a base model with 0.355 accuracy.
- Scale train data and fit the scaler on test & validation data
- Built Logistic regression ,decision trees, xgboost & KNeighbors' models

The Process

✓ *Model Evaluation and Selection:*

- Used 5 folds cross validation against train dataset to evaluate the models.
- Decision trees has the highest mean score for recall since we care about predicting cancelled tenders.

Model	Accuracy	Recall	F1
Logestic Regression	0.646	0.626	0.622
Descision Tree	0.733	0.657	0.664
KNeighbors	0.669	0.643	0.648
XGBClassifier	0.768	0.646	0.649

- Using RandomizedSearchCV to fine tune the Decision tree model and the recall score was 0.66 on test data and validation data.

Tools

- *Sklearn*
- seaborn
- PIL
- wordcloud
- matplotlib

Conclusion

- *Decision trees reached 0.66 recall score in predicting canceled tenders* which means that this model can **save 3300 – ~40K man-hour per month** which will eventually save money on the long run, and this saving is only from the authority side and that will increase if we calculate the saving of cost from vendors & procurement side.
- Based on the model **having no estimated value** is the most important feature in predicting cancelled tenders then **having less than 50 invitations** then **categories count** and having **general tender type**.
- The model can be applied to **prioritize** tenders processing or **highlight** tenders that have high chance of cancellation to procurement department in government sectors.

