FlixCharter: Analyzing the value of bus charter requests

Luis Fernandez, Marit Sloof,
Aurangzeb A. Rathore, Oksana Bilous

February 17th, 2020

FLIXBUS charter

Business Process

- As soon as the customer reserves his or her trip, Charter will split up to the request in two blocks, i.e. into the different trips.
- Charter has two possibilities for assigning a bus to a trip:

Sourcing 1:1

Rental Order

The sourcing team tries to directly find a bus partner that will take care of the trip.

The trip is kept in the optimization pool, for which the Trip Bundler tries to create tours using rental busses (busses rented far in advance).

Glossary

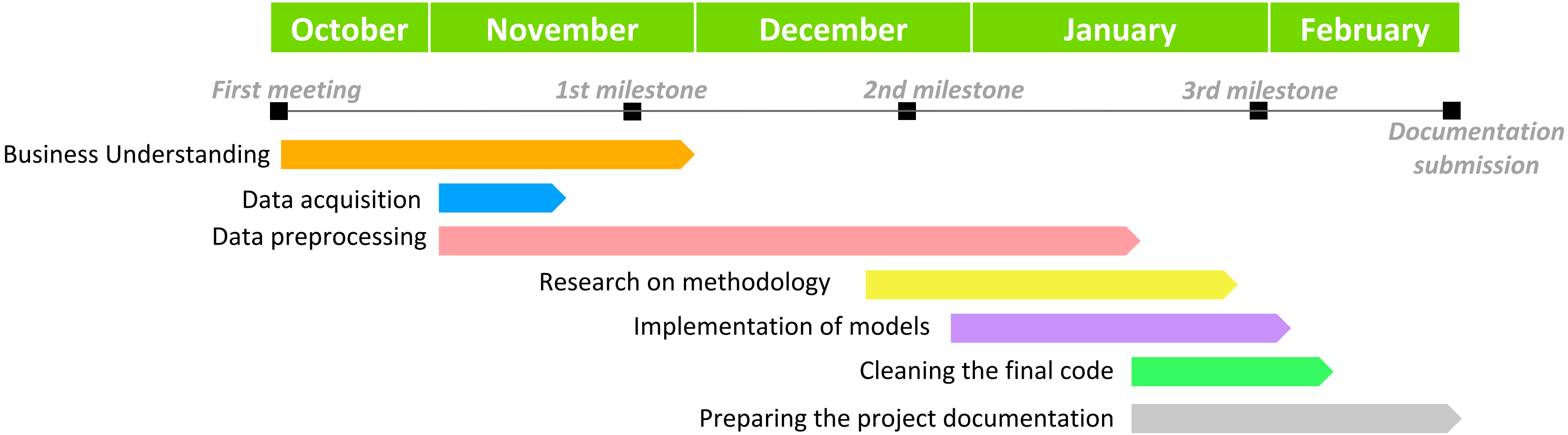
- Trip Bundler: A tool implemented by FlixCharter that creates a chain of trips for a single bus in order to minimize empty kilometers.
- Customer Request (CR): Unique identifier for a customer order. Can contain one or more trips related to it.
- Partner Offer/Sourcing 1:1: When a trip gets handed over to a bus partner.
- Rental Order: When a trip is carried out by FlixCharter using buses they rented in advance.
- Trip Evaluator: The new tool that will serve as a guidance to decide if a trip should be sourced 1:1 or assigned with a rental order.

Why Trip Evaluator?

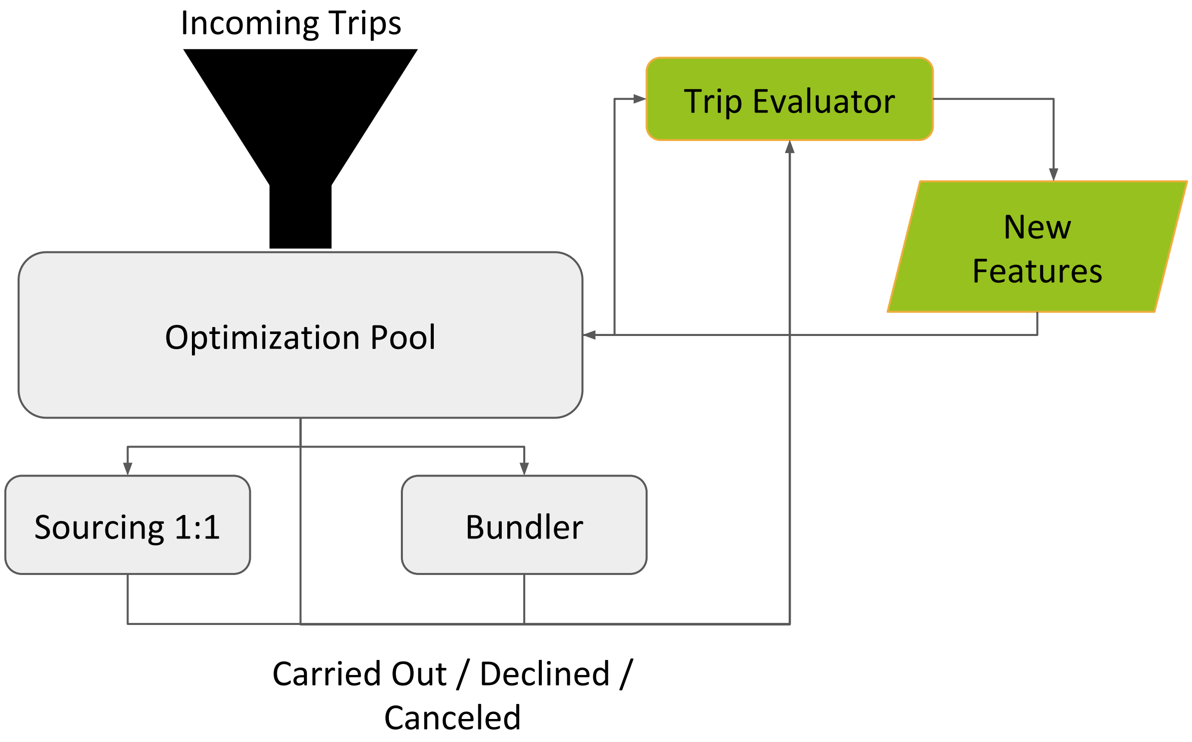
The trip Evaluator attempts to...

- Give a better overview of the trips that could be bundled.
- Anticipate if a trip should be sourced 1:1 or if it should be kept for bundling.
- Predict in advance if a trip could be canceled/declined.
- Inquire whether new trips similar to other are likely to arrive, so we could replace them.
- Estimate how late are we still going to be able to source a trip 1:1 without representing a loss.
- Analyse whether a trip would be hard to source 1:1 and in that case aim to bundle it.

Project Timeline



How does it work?



Cancelation/Declination Probability

Plurality

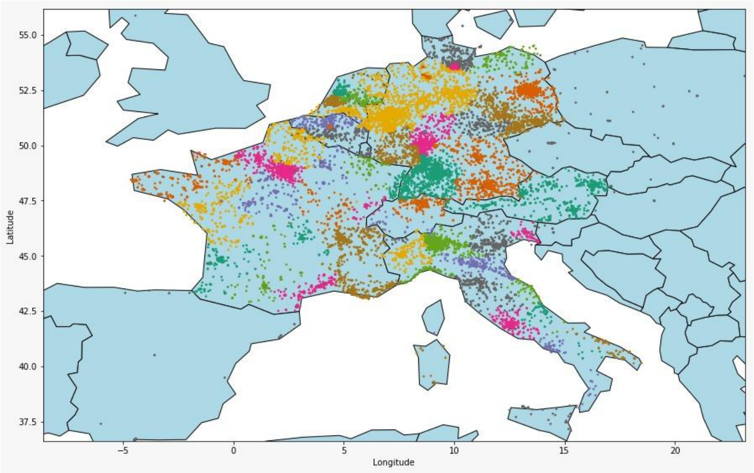
Change Probability

Last/Best Time to Source

Sourcing Complexity

Data

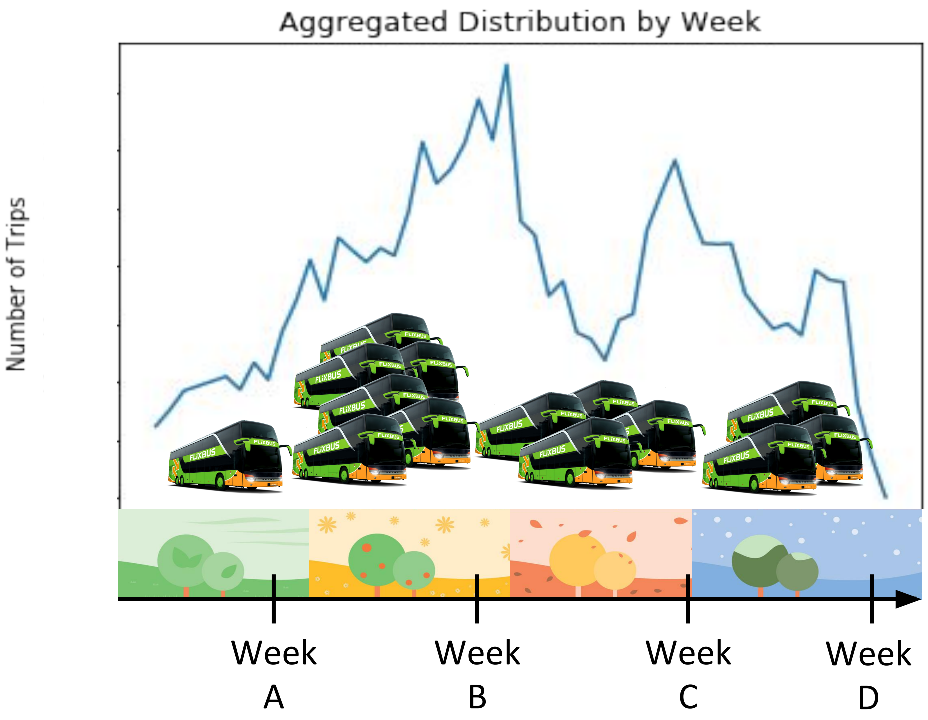
- **Customer request level**
  - Customer type, Trip purpose, price, special requirements, when a customer request changed to a different status, etc.
- **Trip level**
  - Departure/Arrival dates, long/lat, number of passengers.
  - Which requests were assigned with rental orders.
- **Partner Offer**
  - When was a partner offer request created.
  - Number of partner requests, offers.
  - Depots long/lat of the partners.

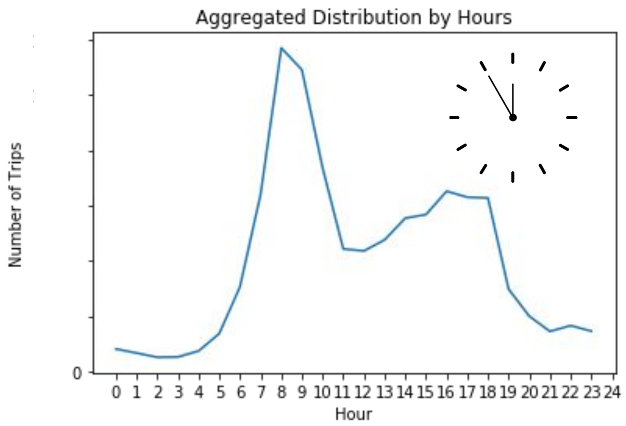


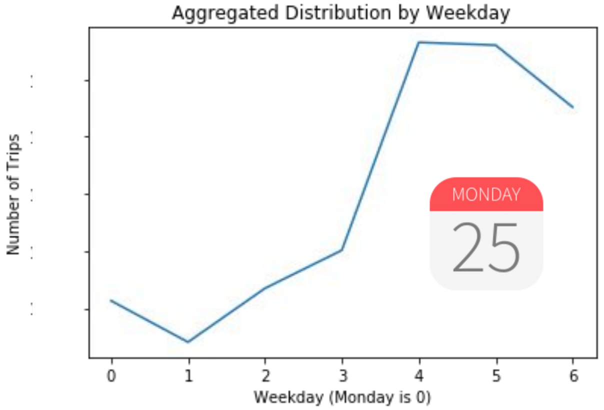
Plurality Calculation

- Measures how common is a trip, in order to know if it's likely to happen again in the future.
- This is calculated distance and temporal wise.

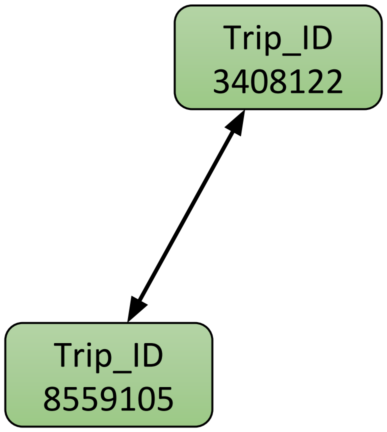
Plurality: Temporal Similarity



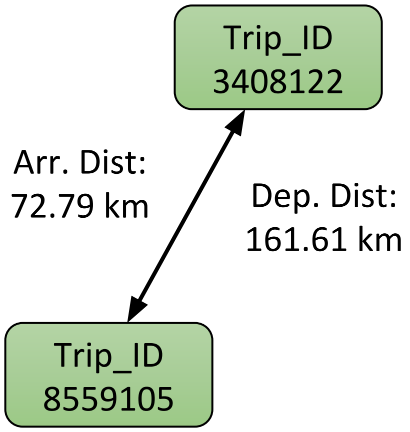




Plurality: Distance Similarity

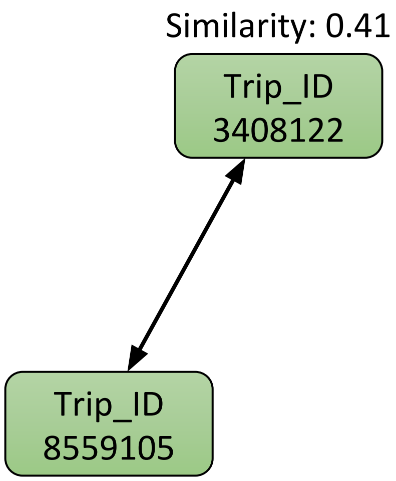


Plurality: Distance Similarity

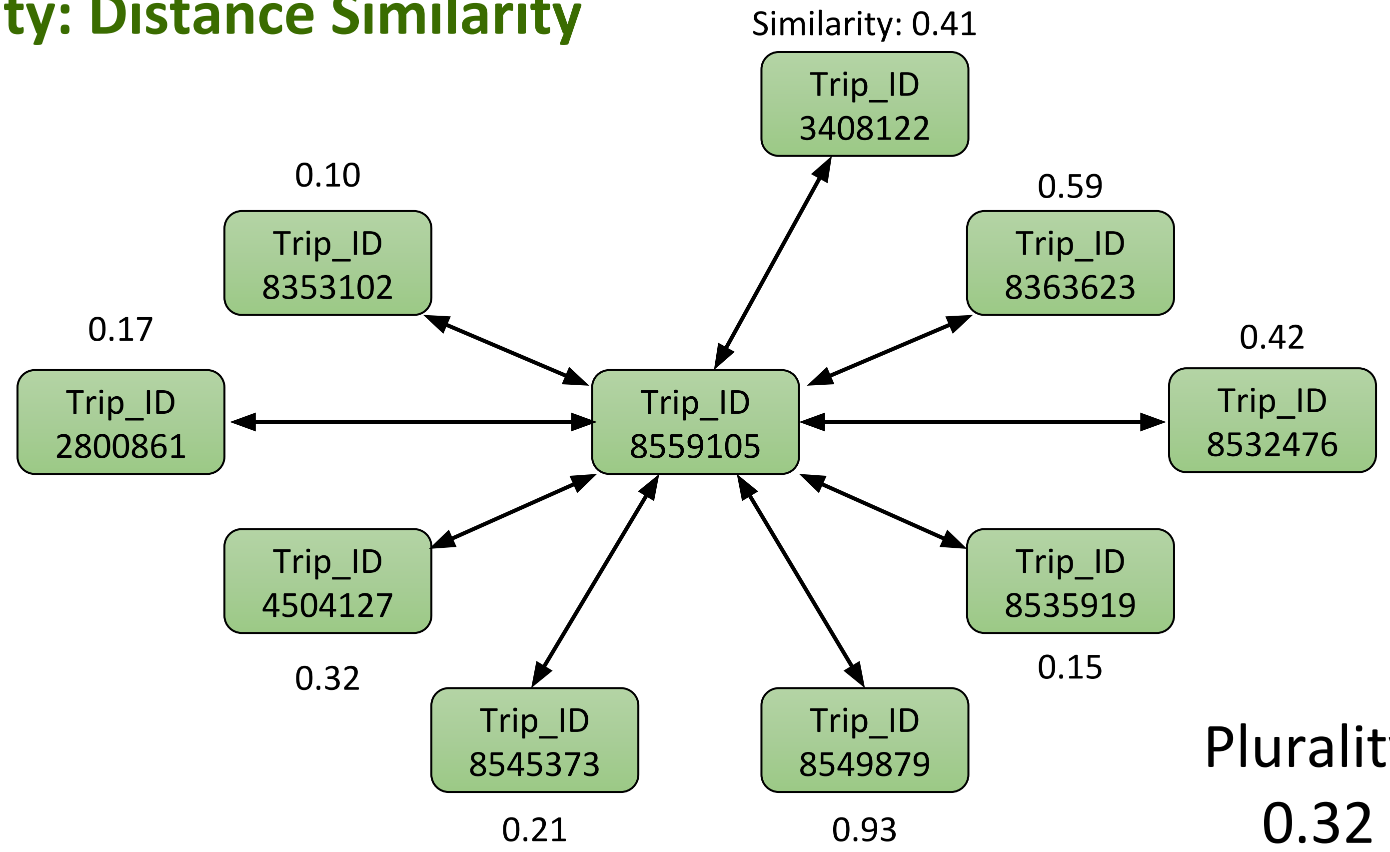


$$Similarity = \frac{Distance(x_{departure}, y_{departure}) + Distance(x_{arrival}, y_{arrival})}{2dMax}$$

Plurality: Distance Similarity

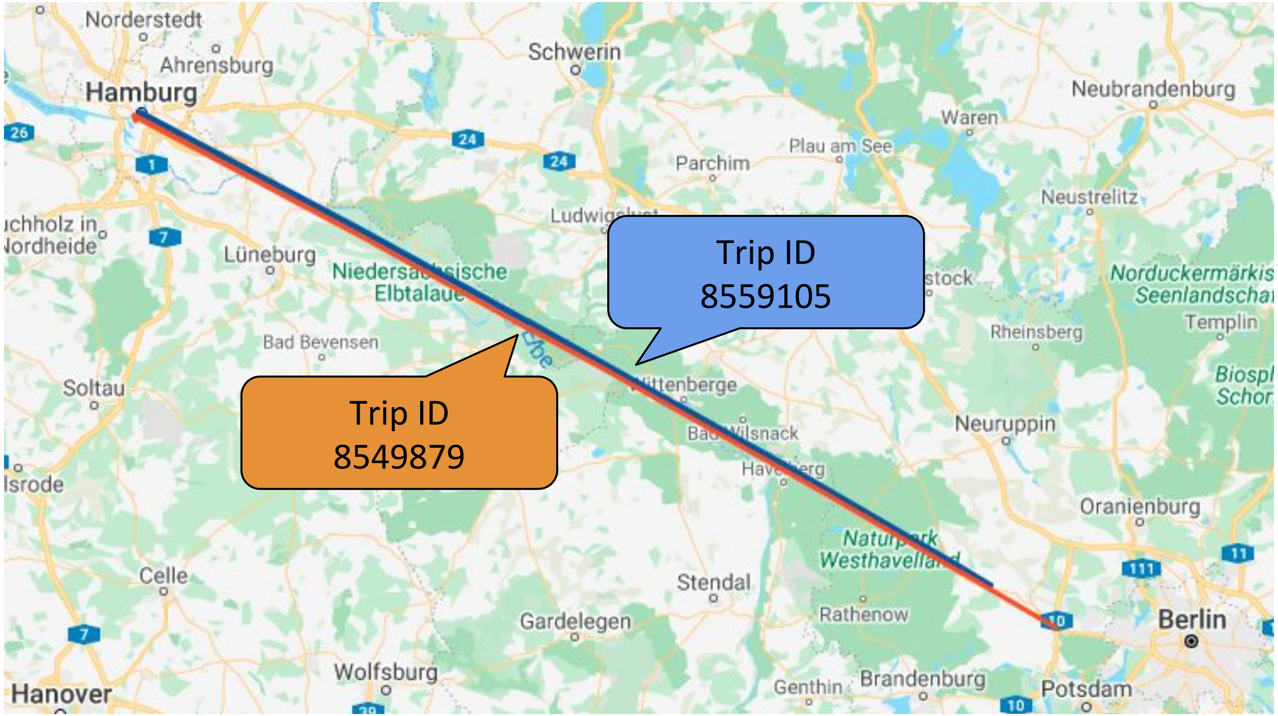


Plurality: Distance Similarity

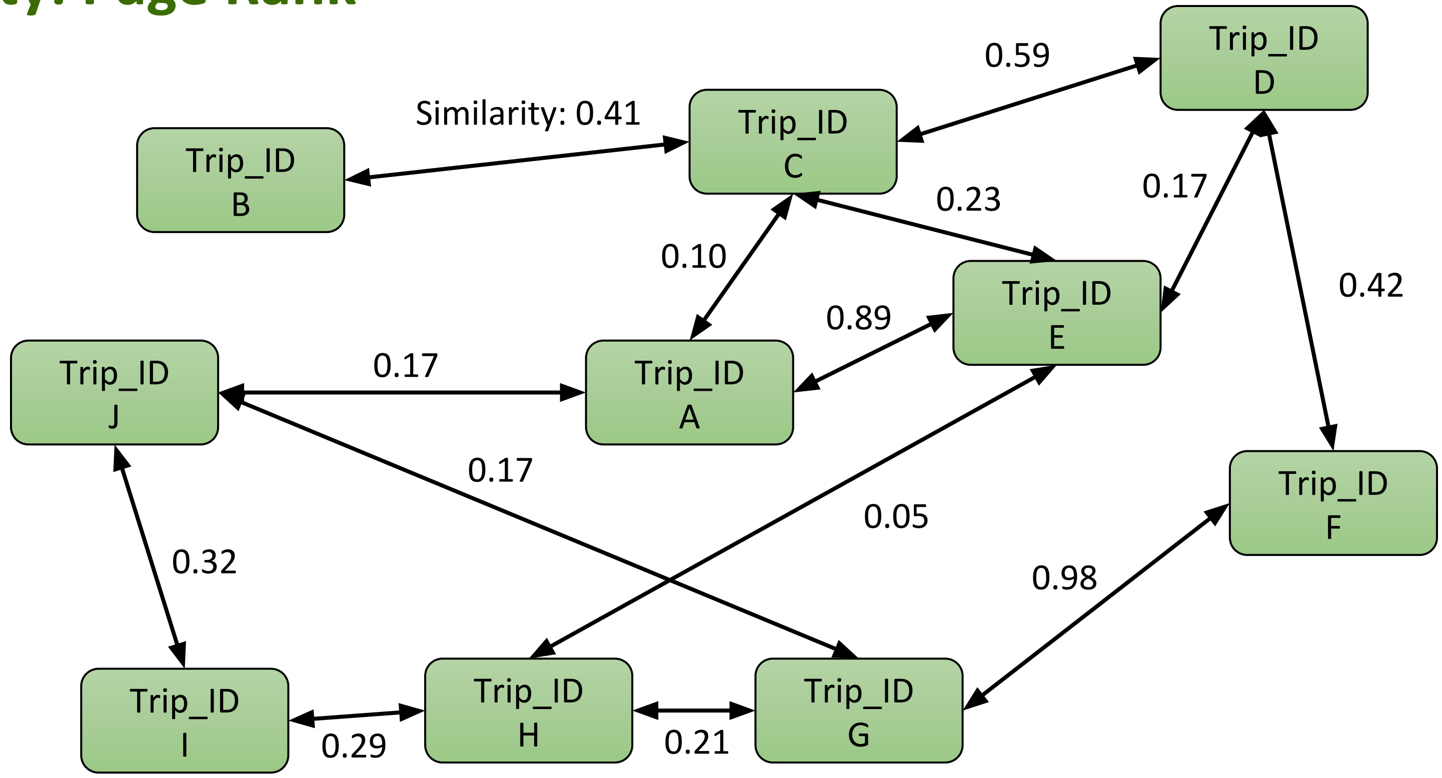


Plurality:
0.32

Plurality: Distance Similarity



Plurality: Page Rank



Plurality: Sparse Matrix

Tested Data

		26918	12678	12387	30821		73981	12639
		0	1	2	3		n +1	n + 2
26918	0		0.13					
12678	1					…	0.16	
12387	2				0.65			
30821	3	0.28						0.32
			…			…	…	
73981	n + 1		0.95					0.32
12639	n + 2				0.73	…		

Tested Data

Plurality: Page Rank Results

Index	Page Rank
0	0.12
1	0.99
2	0.87
3	0.07
4	0.14
…	
n + 1	0.32
n + 2	0.17

Tested Data

Removing outliers and rescaling

Index	Page Rank
0	0.15
1	1.00
2	0.92
3	0.09
4	0.17
…	
n + 1	0.37
n + 2	0.21

73981	0.37
12639	0.21

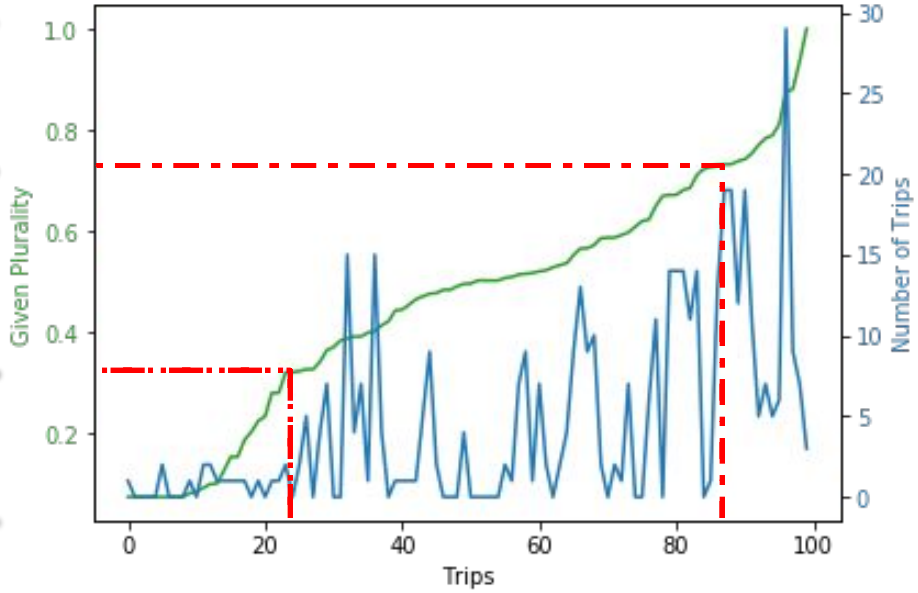
Output

Plurality: Visualizing some Results

New trips are likely

Uncertain about arrival of new trips

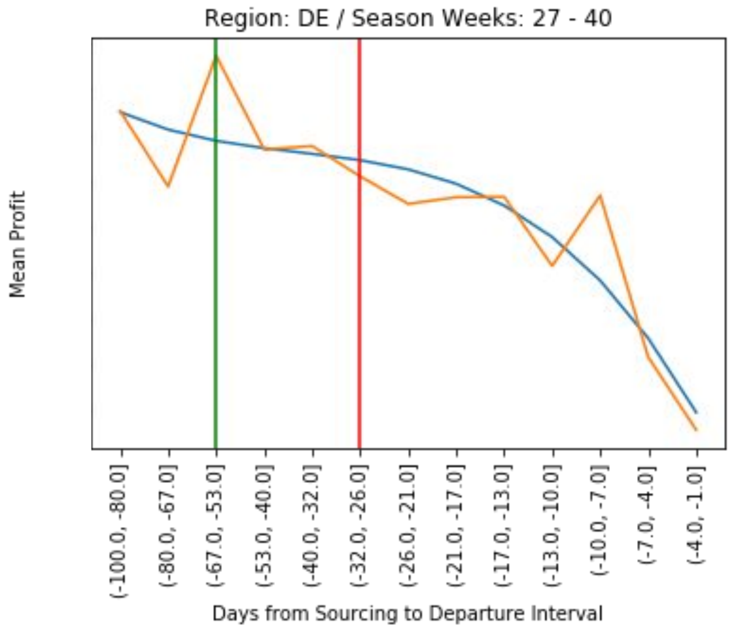
New trips are unlikely

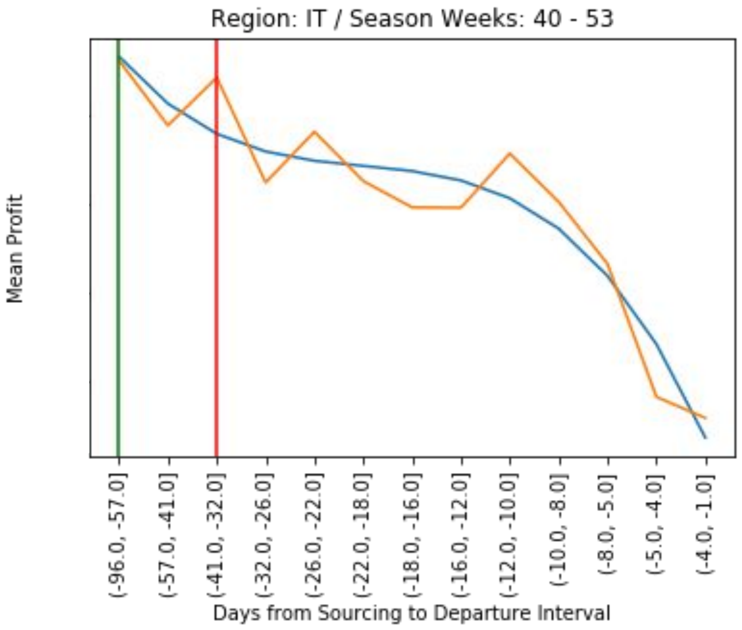


Last/Best Time to Source

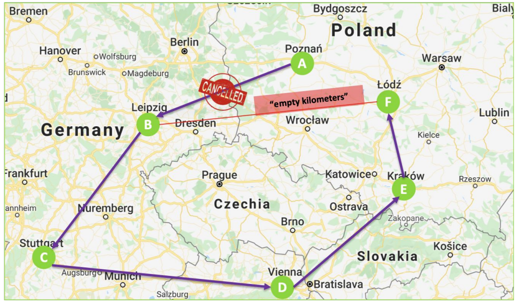
- Aggregate similar trips (Countries and Seasons)
- Find patterns in the mean profit obtained on those trips
- Provide an approximate for best time period for sourcing (highest expected profit) and last time period for sourcing (Time where the profit has reduced but still acceptable)

Last/Best Time to Source





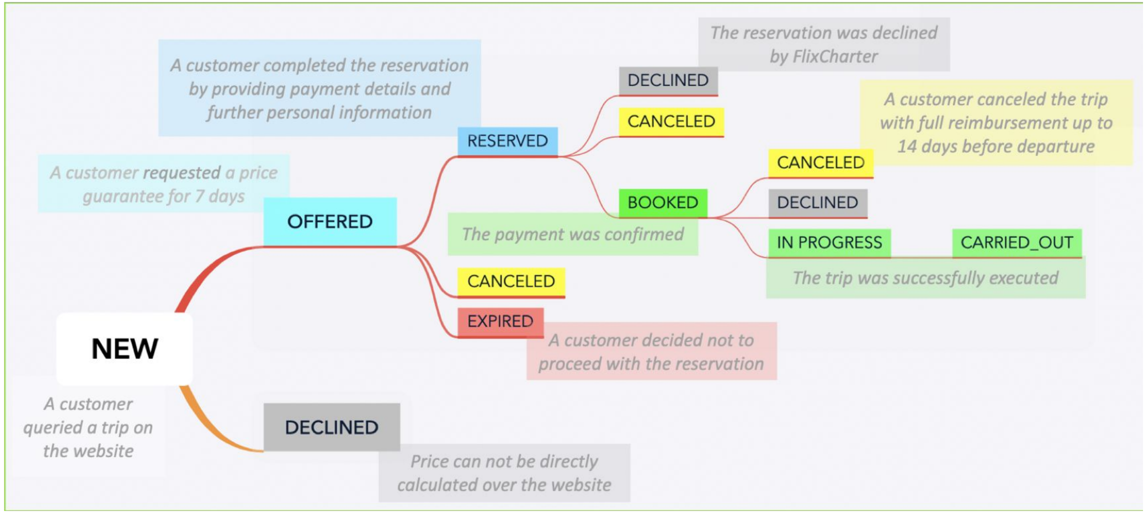
CANC/DECL Probability → Motivation



→ How likely is it for a certain customer request to be canceled or declined?

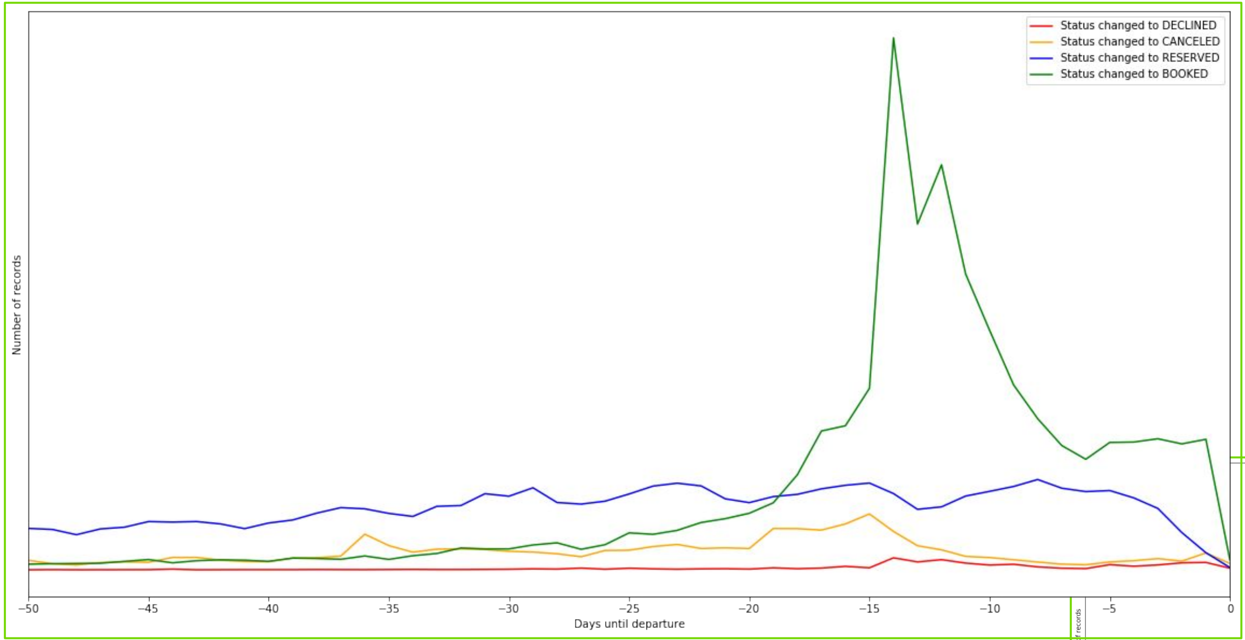
→ Motivation for cancellation / declination probability:

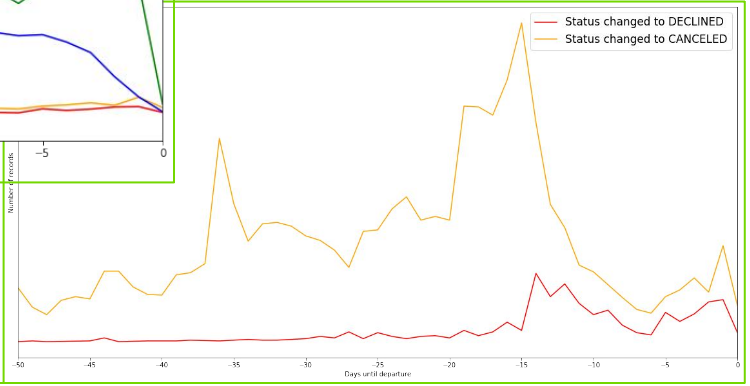
- In most cases, it is not obvious whether a certain customer request will be successfully executed in future.
- The Bundler should not rely too heavy on trips which are rare and are likely to be canceled.
- If such trips were included into a route and later canceled/declined, it could result in a loss for FlixCharter.



- Trips are included into the optimization pool at the time the status is changed to RESERVED
- It is only interesting to predict whether a trip will take place or not
- Classification task with two classes:
  - Class 1: CANCELED/DECLINED
  - Class 0: CARRIED_OUT

CANC/DECL Probability → Analysis of status changes





CANC/DECL Probability → Features used by the model

Categorical attributes

- Customer request type (Instant offer, Website)
- Customer origin (Website, Mail, Chat, Phone)
- Non standard flag
- Voucher redemption flag
- Language spoken in departure/arrival country flag
- Customer type (Club, School, Individual, etc.)
- Trip purpose (School trip, Club trip, Day trip etc.)
- Payment type (On account, SEPA)
- Luggage request (Normal, Large)
- Weekday of departure
- Season of departure
- Departure country (DE, FR, IT)

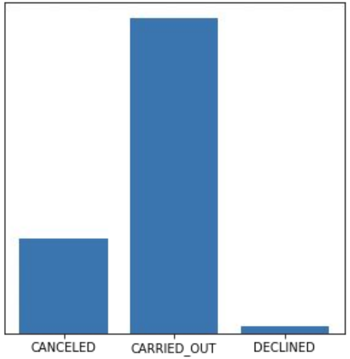
Numerical attributes

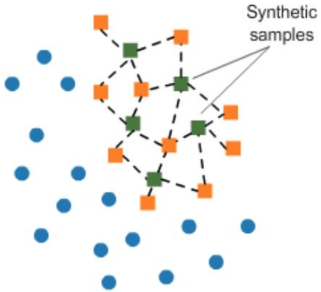
- Number of buses
- Number of trips
- Number of passengers
- Number of special requirements
- Total distance in km
- Longitude & latitude of departure point
- Duration and number of breaks
- Price rate of the market price
- Year of departure
- Days from reserved to departure
- Days from 01.01.2017 to departure
- Number of cancellations/declinations and number of executed trips for this customer

CANC/DECL Probability → Imbalanced classes

- Random Oversampling
  - Naively creating random copies of the minority class.

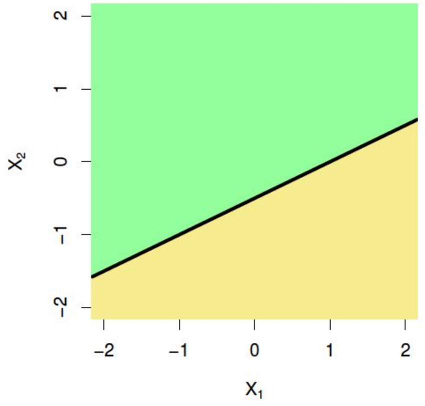
- Synthetic Minority Oversampling TEchnique (SMOTE)
  - Introducing a synthetic observation with every minority observation along the line segments that connect to the nearest surrounding neighbors.





CANC/DECL Probability → Logistic Regression Classifier

- Models probabilities in linear functions of the explanatory variables.
- Advantages:
  - Measures marginal contribution per explanatory variable.
- Disadvantages:
  - Usually performs worse on accuracy.

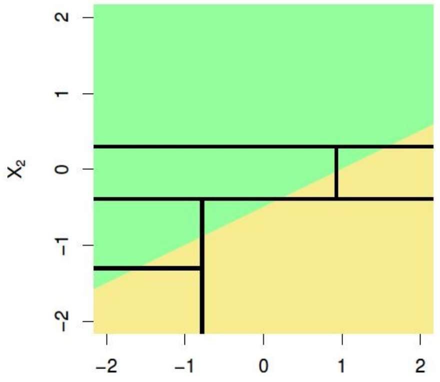


CANC/DECL Probability → Random Forest Classifier

- Decision tree ensemble method.

- Advantages:
    - Generally scores high on accuracy.
    - Overall importance of features can be measured by the Gini index.

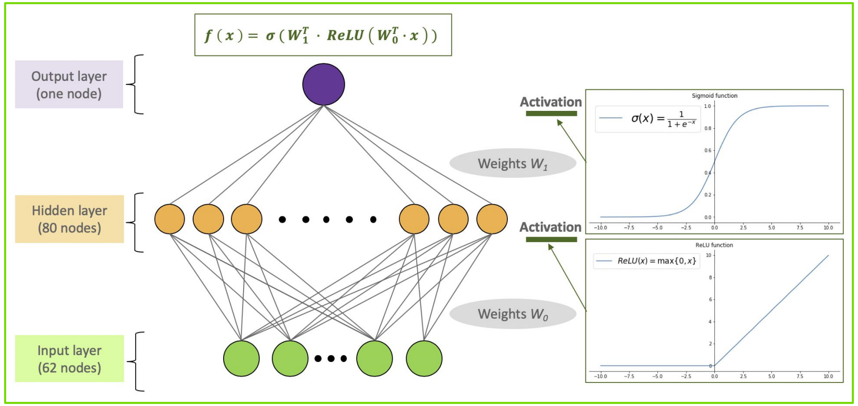
- Disadvantages:
    - Ensemble of trees makes the results difficult to interpret.



CANC/DECL Probability → Neural Network Classifier

- Feed-Forward Single-Layer Neural Network
- Input: normalized customer request matrix $\widehat{\boldsymbol{X}} \in [0,1]^{N \times d}$ and true labels $\boldsymbol{y} \in \{0,1\}^N$
- Rectifier for the first activation function and sigmoid activation for the output layer
- Binary cross-entropy between $\boldsymbol{f}(\boldsymbol{x_i}, \boldsymbol{W})$ and the true labels $\boldsymbol{y_i}$ as the cost function for backpropagation

$$E(W) = -\sum_{i=1}^{N} y_i \log f(x_i, W) + (1 - y_i) \log(1 - f(x_i, W)))$$



Canc/Decl Probability → Estimation metrics for hyperparameter optimization and model evaluation

→ How can we evaluate the performance of different models?

First idea: **Accuracy** which indicates the percentage of records that were classified correctly:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}$$

→ Drawback: Due to imbalanced dataset (proportion of classes is 3:1) accuracy might be high just by classifying the CARRIED_OUT CRs correctly and failing to identity CANCELED/DECLINED CRs.

Second idea: having a look at the **confusion matrix**

- **Recall** of class k indicates what percentage of records in class k were correctly classified:

$$r_k = \frac{TP}{TP+FN}$$

- **Precision** of class k indicates what percentage of records classified as class k really belong to class k:

$$p_k = \frac{TP}{TP+FP}$$

- **F1 Score** is the harmonic mean of recall and precision:

$$F1 = \frac{2 * Recall * Precision}{Recall + Precision}$$

- **AUC** (Area Under Curve) represents a measure of separability. It tells how much model is capable of distinguishing between classes. The AUROC tends towards 1.0 for the best case and towards 0.5 for the worst case.

	Predicted label CARRIED_OUT	Predicted label CANCELED / DECLINED
True label CARRIED_OUT	TN	FP
True label CANCELED / DECLINED	FN	TP

CANC/DECL Probability → Hyperparameter optimization

- Logistic Regression Classifier:
  - Feature forward selection and significance of parameters

- Random Forest Classifier:
  - Random grid search over the hyperparameters (such as number of estimators, min. samples split, max. depth etc.) with 5-fold cross validation using *GridSearchCV* from the *ScikitLearn* library.

- Neural Network Classifier:
  - Training a neural network with different hyperparameters (batch size, number of epochs, layers and number of nodes in those layers), then comparing estimation metrics for all sets of hyperparameters

CANC/DECL Probability → Model evaluation

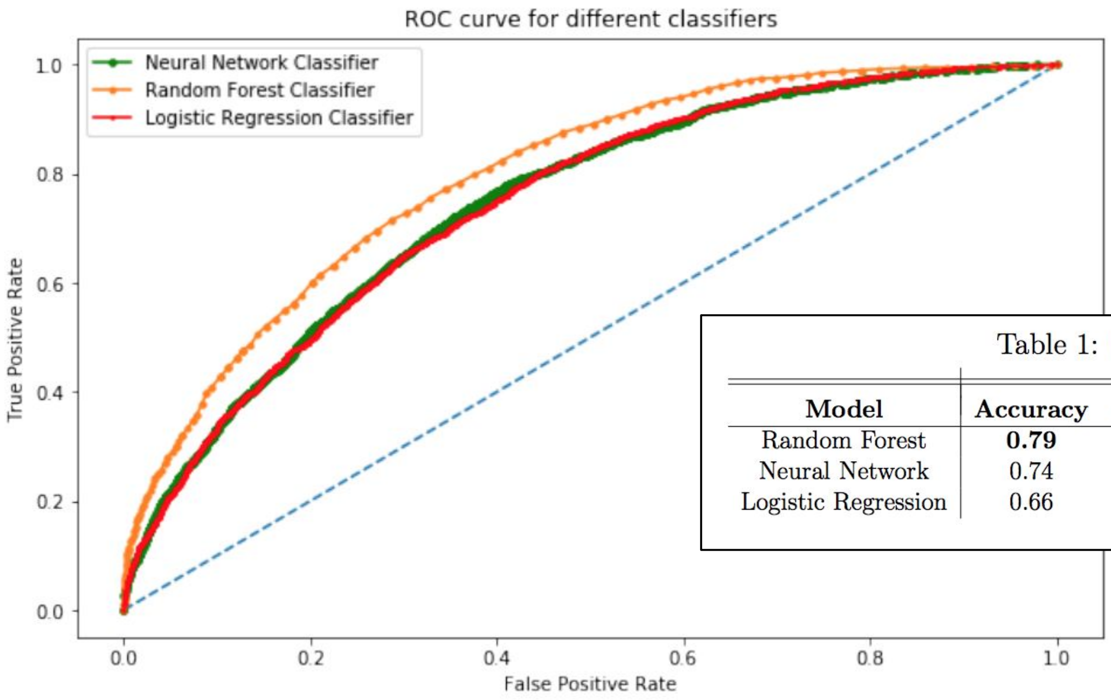
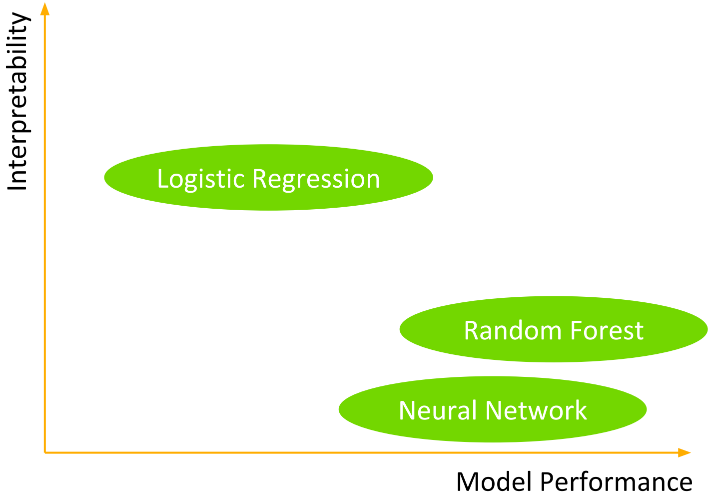


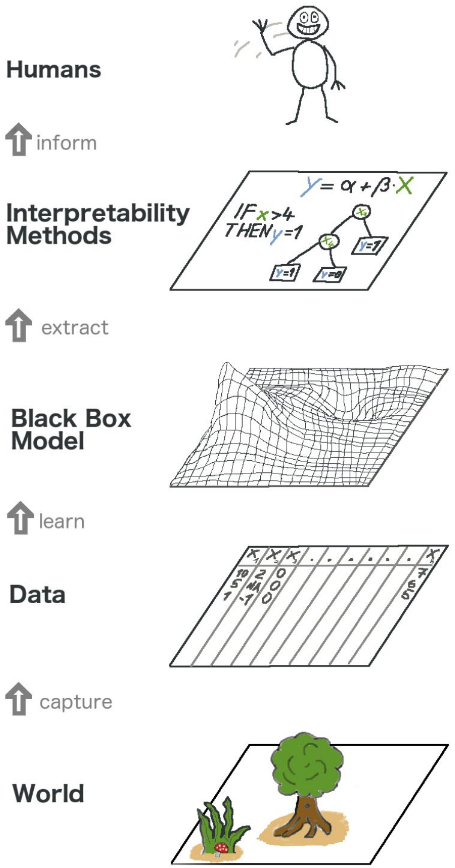
Table 1: Model results

Model	Accuracy	AUC	Precision	Recall	F1 score
Random Forest	**0.79**	**0.8**	**0.57**	0.44	**0.5**
Neural Network	0.74	0.75	0.46	**0.45**	0.45
Logistic Regression	0.66	0.62	0.37	0.56	0.45

CANC/DECL Probability → Model interpretability

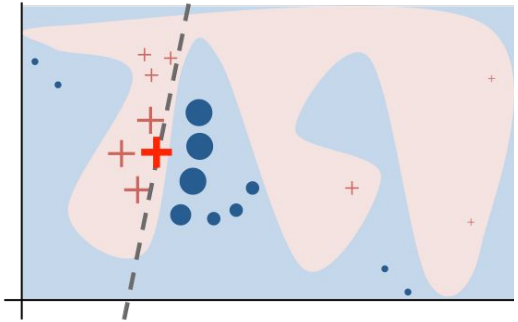
- Interpretability of powerful "black box" classifiers is essential for establishing trust of users in the predictions produced by the model
- We are looking for an approach that can reveal the relationship between input and output without having a thorough understanding of a specific model



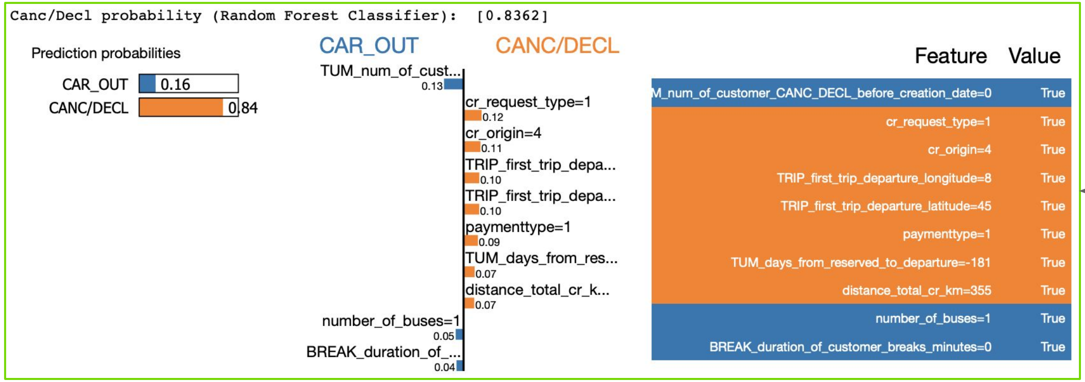


CANC/DECL probability → Model interpretability → LIME

- **Local Interpretable Model-Agnostic Explanations (LIME)** can be applied to any ”black-box” model.
- The idea is to learn an <u>interpretable</u> model (e.g. a linear model) <u>locally</u> around the prediction.
- The coefficients of the linear decision boundary reveal the impact of the features of a data point on its class prediction
- Python library *lime* was used to produce explanations (see examples on the next slide)

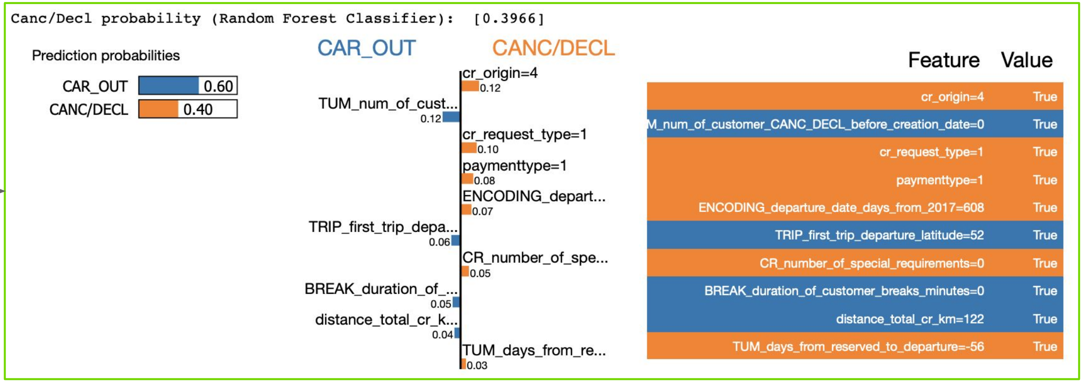


CANC/DECL probability → Model interpretability → LIME: examples

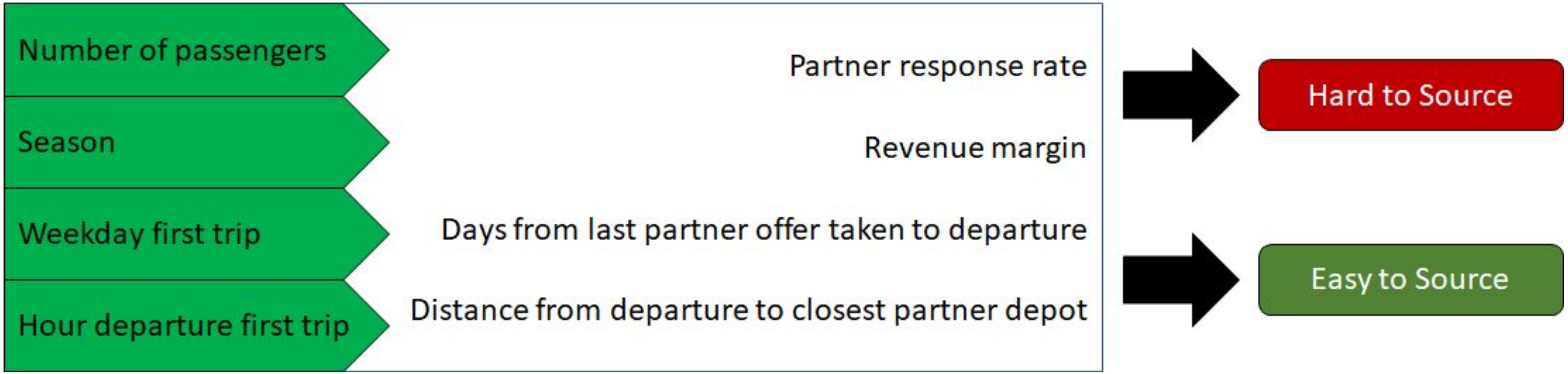


LIME output for a random customer request with the final status CANCELED

LIME output for a random customer request with the final status CARRIED_OUT



Sourcing Complexity (Heuristics)



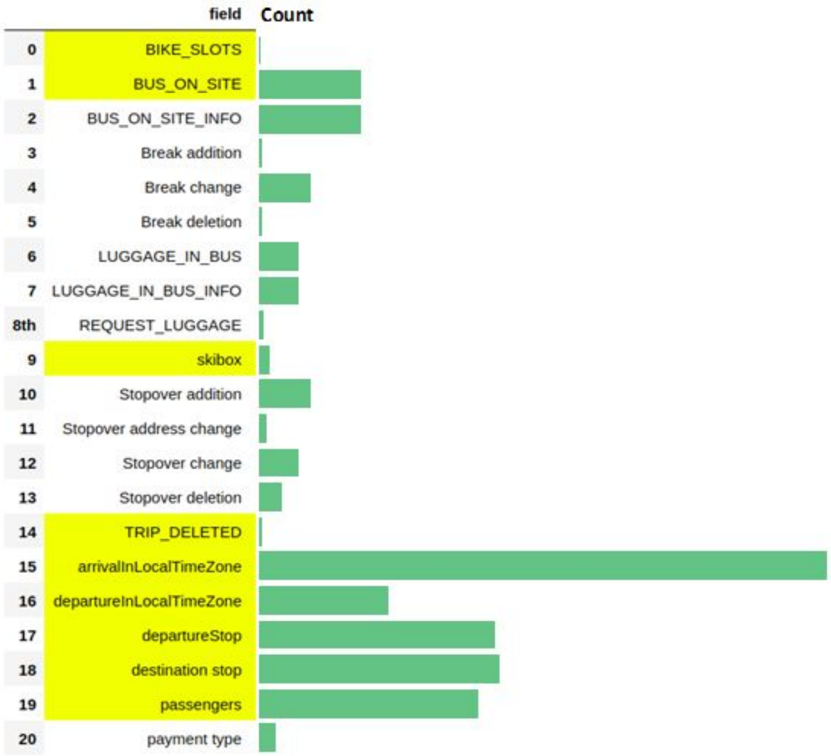
Sourcing Complexity (Clustering)

- A few CRs were labeled as hard/easy to source
- Attempted to cluster on parameters relevant for this feature
- No good results were reached, changed drastically while re-sampling. More samples could improve the accuracy.

Change probability: Columns

- **Id:** Unique identifier for each change record
- **Created:** Date of the change made
- **Field**: Type of change made to the order
- **currentvalue**: The updated value of the change made with respect to the field
- **previousvalue**: Previous value of the record that was changed
- **Customerrequestid:** Identifier of the customer corresponding to the change was made
- **trip_id**: Identifier of the trip corresponding to the change that was made

Fields



BIKE_SLOTS

BUS_ON_SITE

Skibox

TRIP_DELETED

ArrivalLocalTimeZone

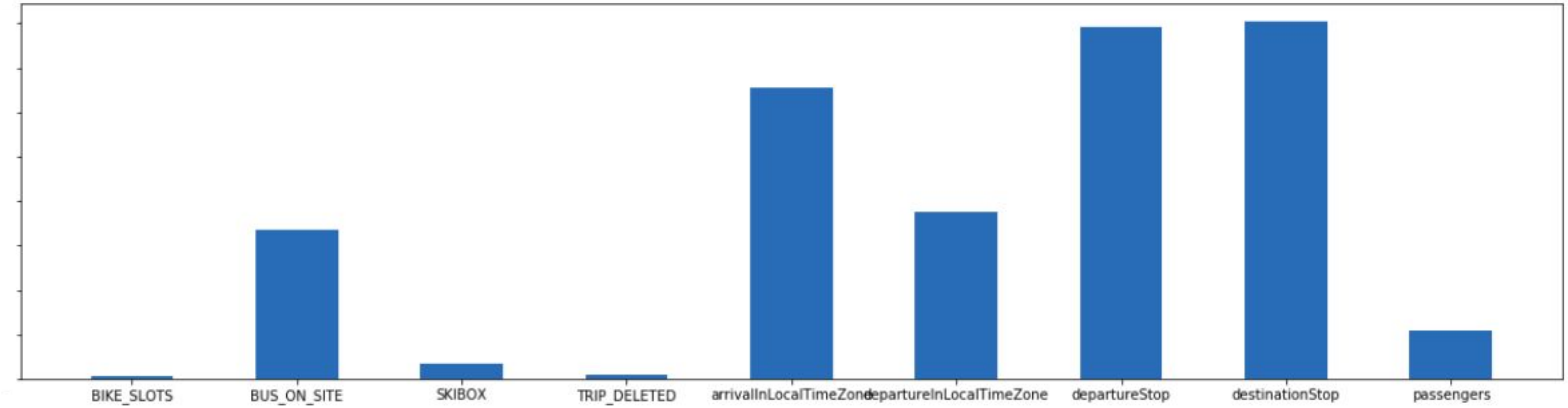
DepartureinLocalTimeZon e

DepartureStop

DestinationStop

Passengers

Fields


BIKE_SLOTS
BUS_ON_SITE
SKIBOX
TRIP_DELETED
arrivalInLocalTimeZone
departureInLocalTimeZone
departureStop
destinationStop
passengers

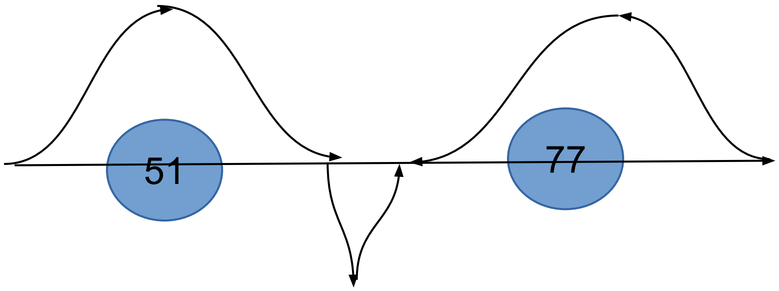
Criteria for Significant Change

- **Skibox** or **Bikeslots** or **Bus_on_site** or **Trip_deleted** == True → Mark as Significant

- **ArrivalLocalTimeZone** or **DepartureLocalTimeZone** change >= |0.5| → Mark as Significant

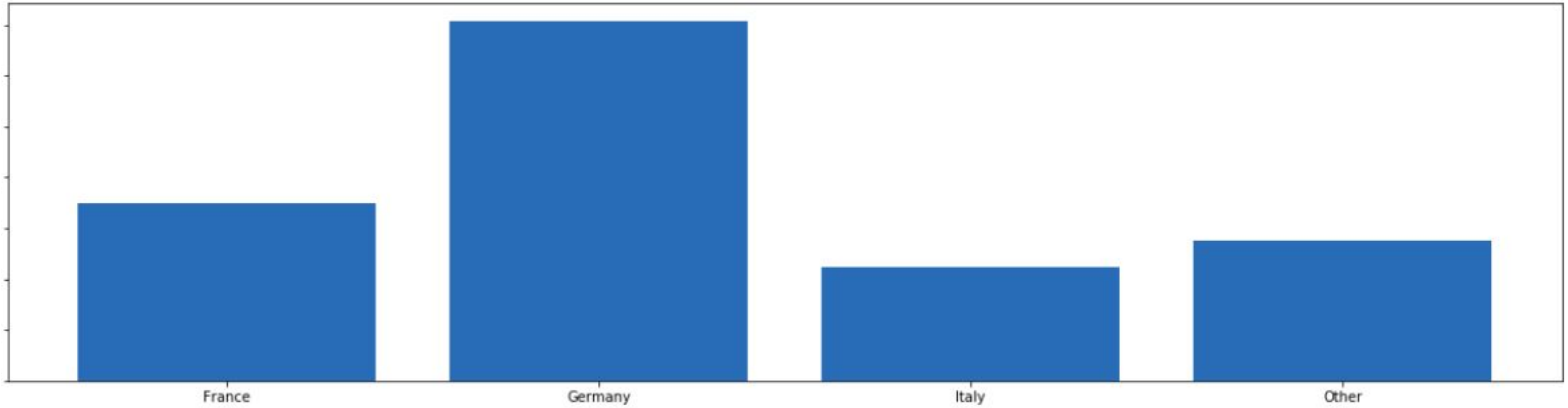
- **DepartureStop** or **DestinationStop** == Change inclined for new distance measure → Mark as Significant

- **Passengers**

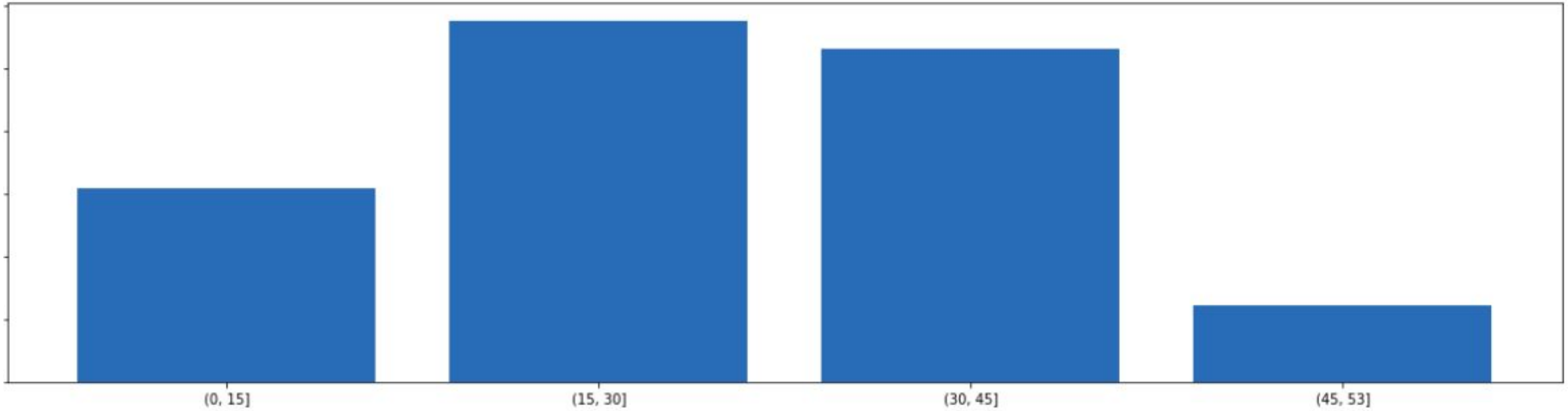


→ Mark as Significant

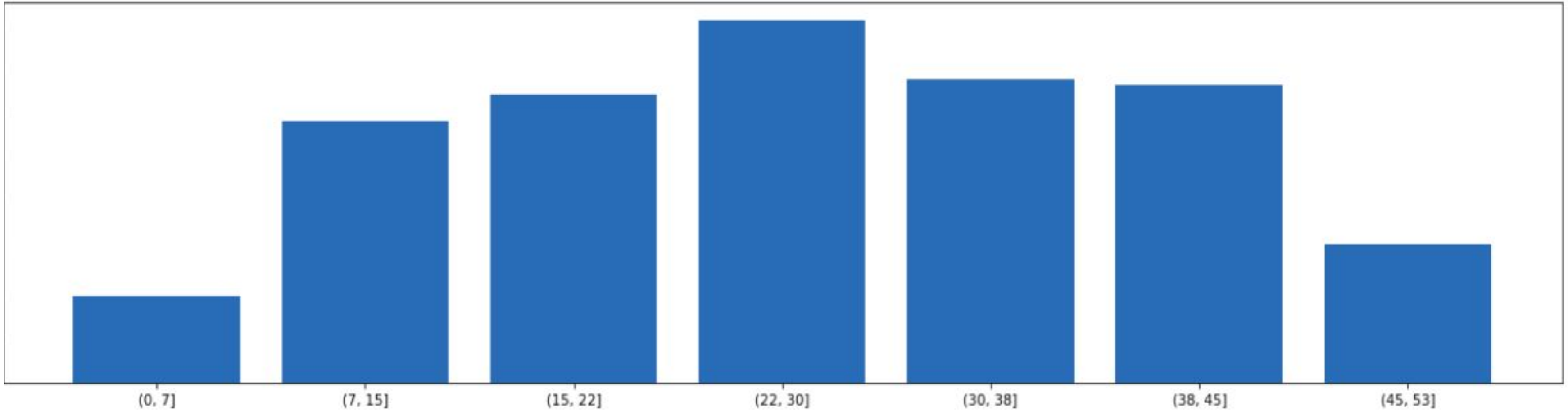
Distribution of Significant changes with respect to Countries



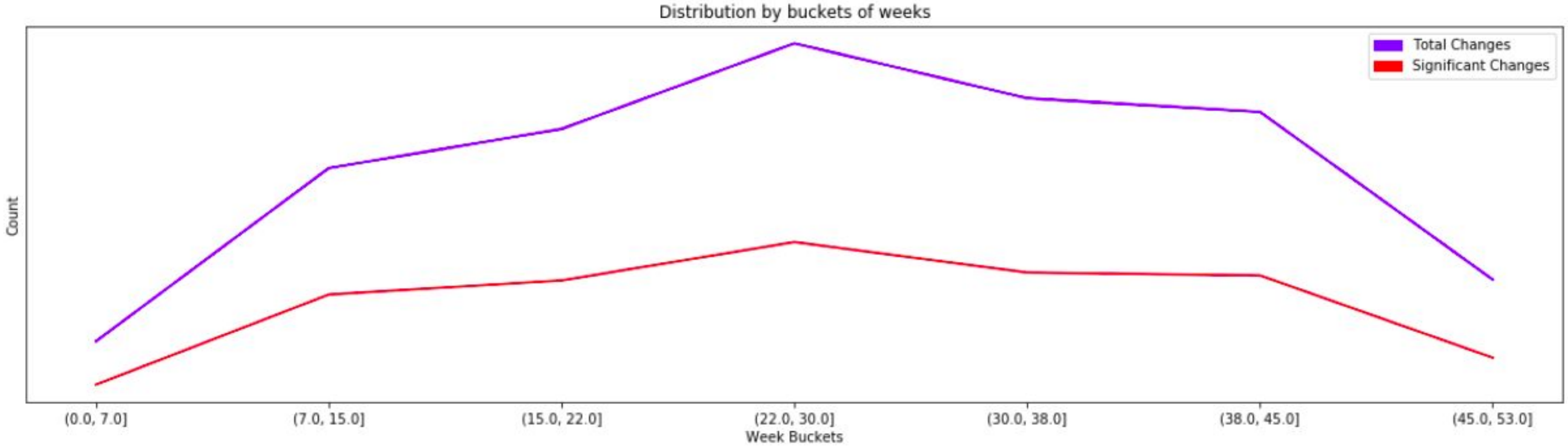
Distribution of Significant Changes with respect to week buckets



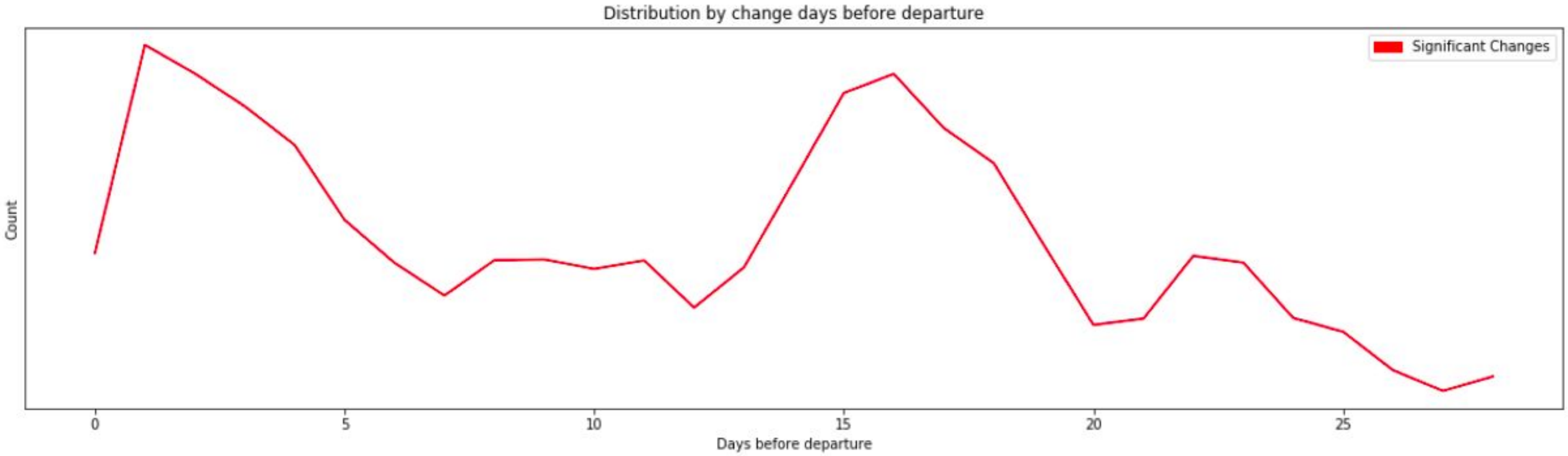
Distribution of Significant Changes with respect to week buckets



Distribution of Significant Changes with respect to week buckets

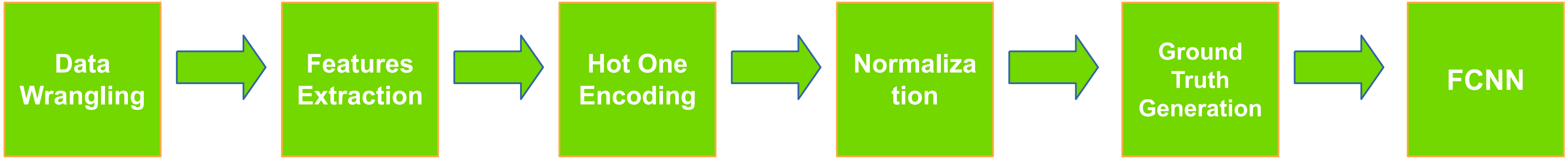


Days before departure vs Count

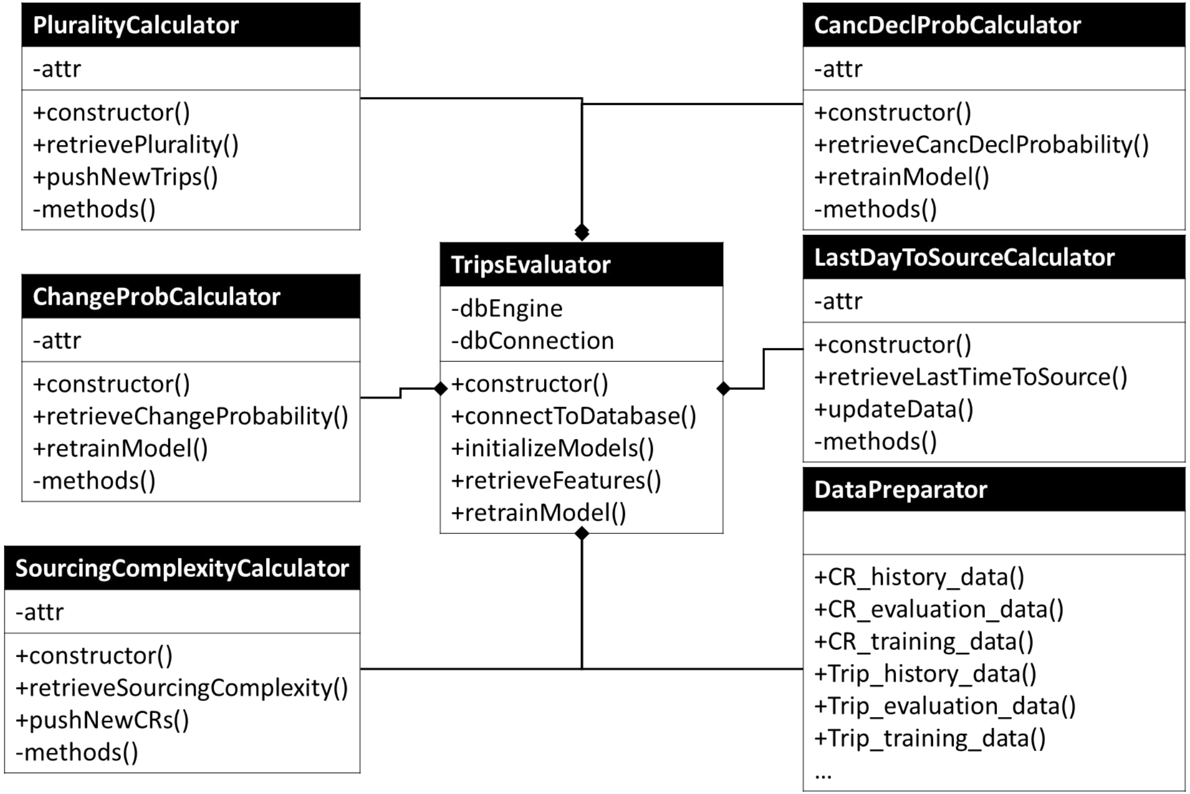


FLiX

Change Process



The current library structure



FLiX

Thank you for your time

Questions? :)