

The background is a dark teal color with various financial symbols and numbers scattered across it. Symbols include the dollar sign (\$), the euro sign (€), the pound sign (£), and the yen sign (¥). Numbers are in various sizes and colors (white, yellow, and teal). Some numbers are accompanied by upward or downward arrows, suggesting trends or fluctuations. The overall theme is financial and data-related.

Credit Card Fraud Detection

Adam Blomfield
Kelly Epley
Helen Levy-Myers

What is Fraud?

FRAUD

An unauthorized transaction

SCAMS

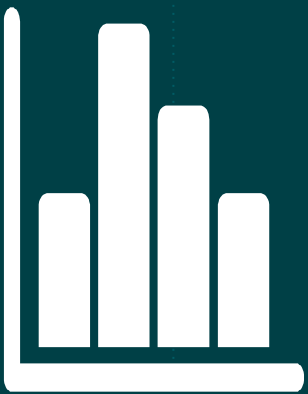
An authorized transaction,
but not as promised



Who Cares About Credit Card Fraud?

- Credit Card Issuers
- Merchants
- Consumers





DATA SET

Credit Card Fraud Data Set From Kaggle

284,807

European credit card transactions from two days in September 2013

492 fraud cases

Making a very unbalanced data set

31 variables

Amount, Time, Fraud or Valid and 28 masked variables through PCA

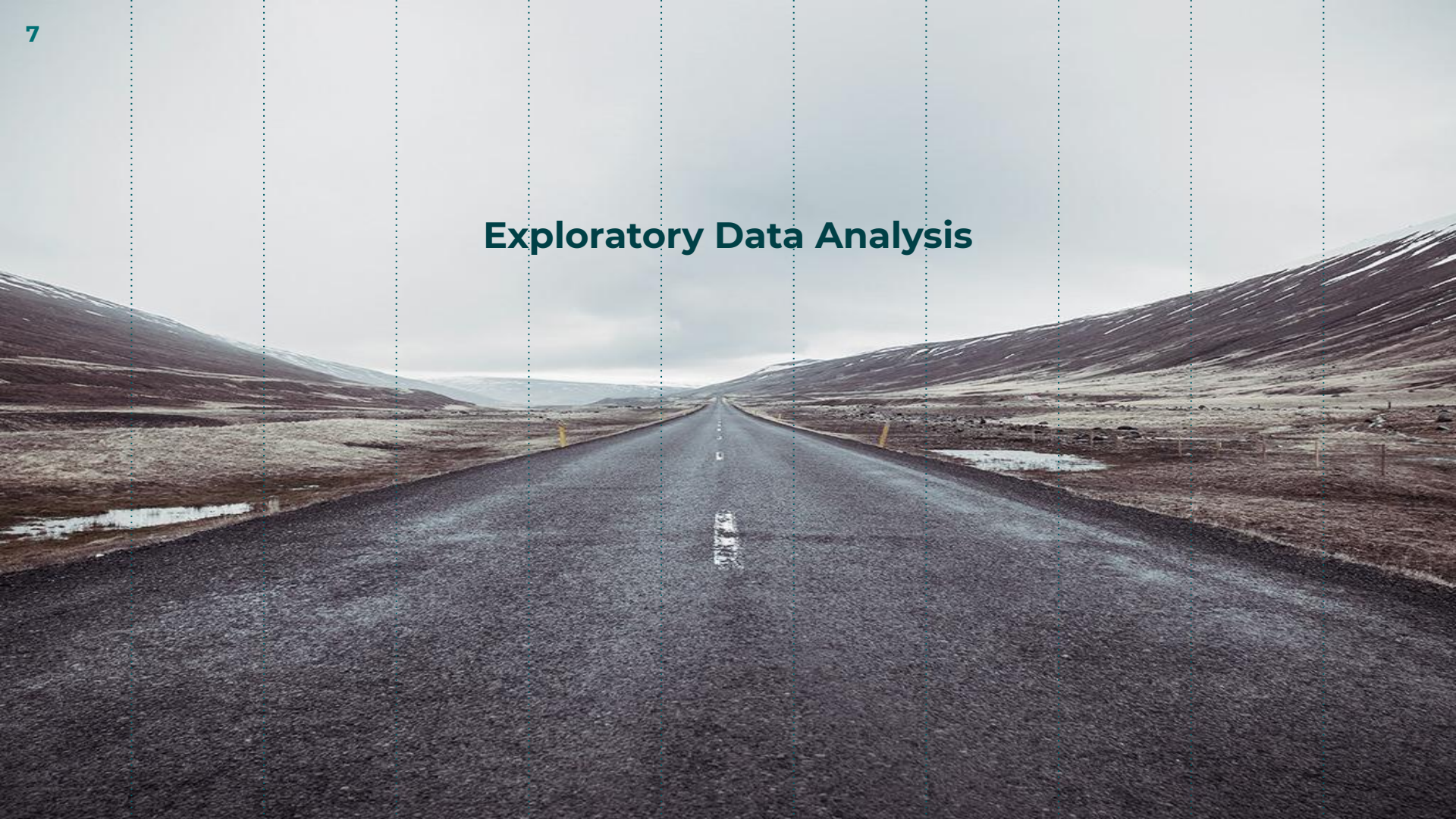
Data Cleaning

Eliminated duplicates

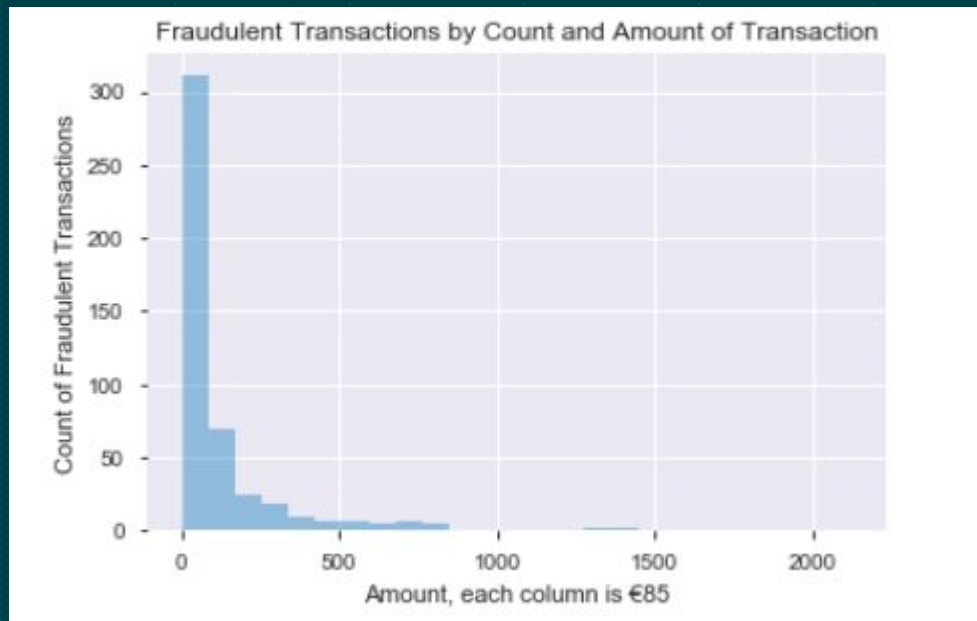
No nulls or missing values



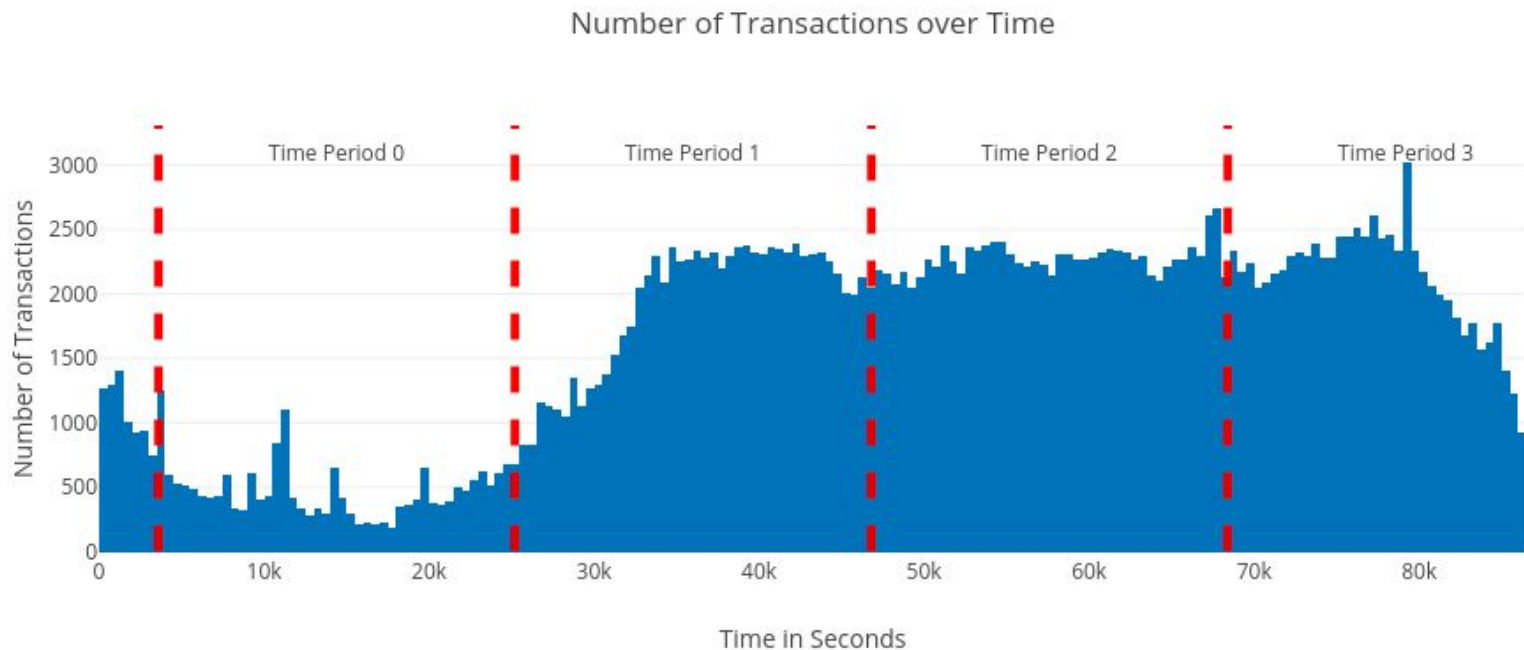
Exploratory Data Analysis



Frauds Skewed to Low Amounts



Credit Card Transactions by Time Period



The Models Not Taken



Models Considered

Logistic Regression

Produced poor results



SVM

Produced poor results



Considering only larger fraud

Reduced fraud data even more
Might introduce new
opportunities for fraud



Model and Sampling Evaluation

Classifier	Oversampled minority	Undersampled majority class	Mix of Both
Random Forest	.63	.14	.27
Logistic Regression	.04	.03	.05
K-nearest Neighbors	.27	.14	.06
Support Vector Machine	.06	.04	.09
Gradient Boosting.	.30	.07	.14

F1 Scores on the Testing Data

F1 Score

A weighted average of precision and recall

Precision

$$\frac{\text{TruePositives}}{\text{TruePositives} + \text{FalsePositives}}$$

Recall

$$\frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}}$$

F1

$$2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

The Final Model



Random Forest

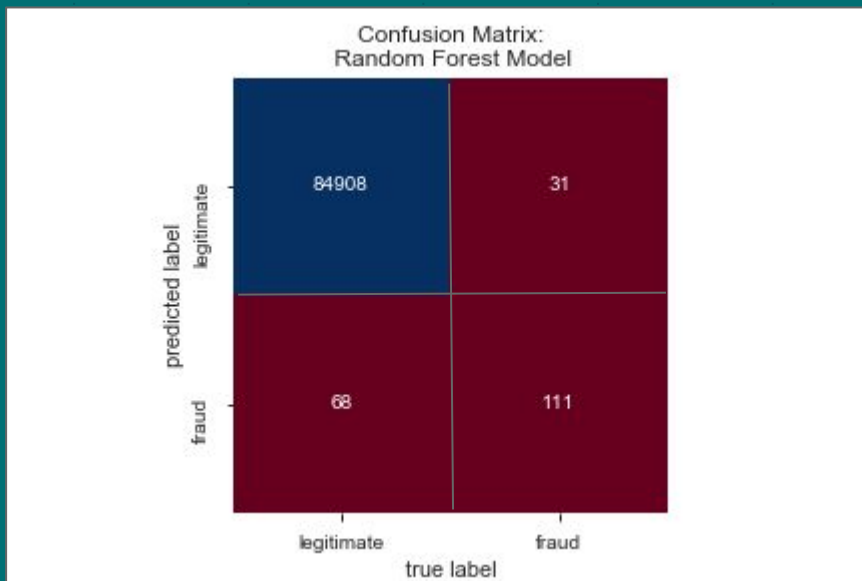


Oversampling with SMOTE

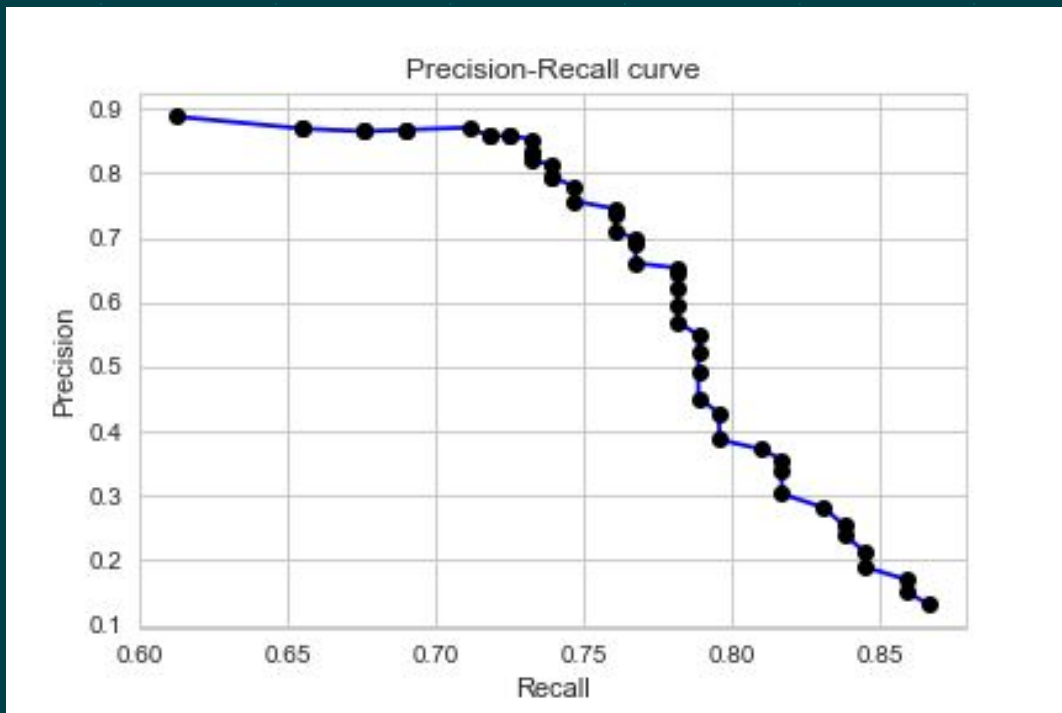
Class Weighting

Grid Search

Confusion Matrix

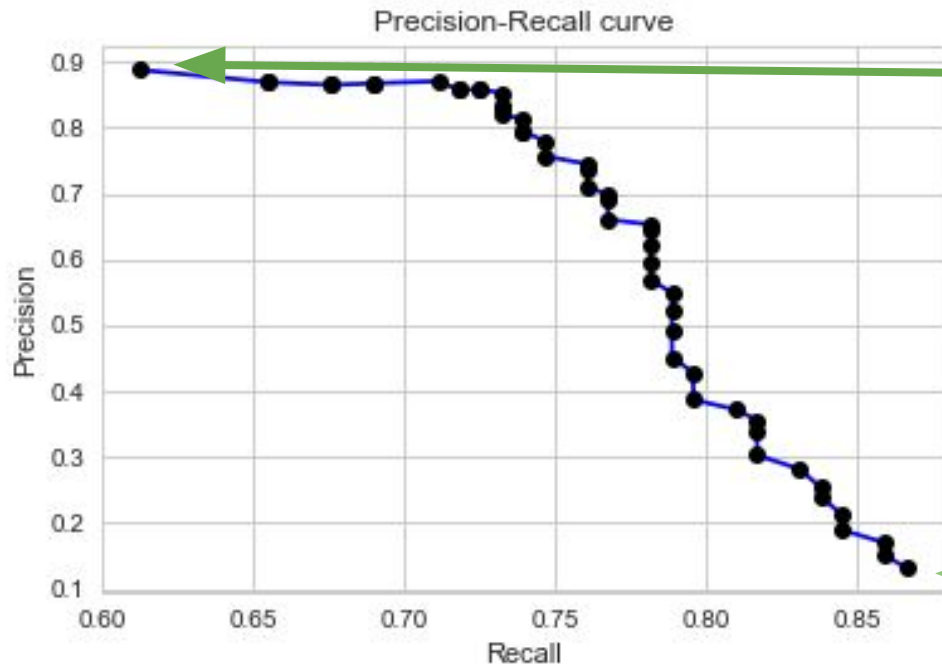


Precision-Recall Curve



Shows how precision & recall change as we adjust the probability threshold for a positive identification

Precision-Recall Curve



High Threshold:

High Precision Means Fewer False Positives, Low Recall Means More False Negatives

Low Threshold:

Low Precision Means More False Positives, High Recall Means Fewer False Negatives

Thanks!

Any questions?

Kelly Epley, kemepely at gmail.com

Adam Blomfield, bcd@xyz.com

Helen Levy-Myers, hlevymyers at gmail.com



Credits

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by [SlidesCarnival](#)
- Photographs by [Unsplash](#)

