



CSE 4062 – Introduction to Data Science and Analytics
Spring 2021
Delivery #2 - Exploring Data

Project Report - Group #8

Group Members:

Caner Dağdaş | 150716001 | canerdagdas@hotmail.com
Ceyhun Vardar | 150317022 | vardarceyhun13@gmail.com
Büşra Gökmen | 150116027 | busragokmen67@gmail.com
Cem Güleç | 150117828 | cem.ggulecc@gmail.com
Ömer Faruk Çakı | 150117821 | omerfarukcaki@gmail.com

Project Title: Fraud Detection on Financial Data

Lecturer: Assoc. Prof. Murat Can Ganiz

There are 11 columns as listed below. Since the dataset is synthetic there is no missing value occurring in any of the attributes.

Feature Names	Description	Type	Missing Value (%)
step	maps a unit of time in the real world	numeric	0
type	CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER	nominal	0
amount	amount of the transaction in local currency	numeric	0
nameOrig	customer who started the transaction	nominal	0
oldbalanceOrg	initial balance before the transaction	numeric	0
newbalanceOrig	new balance after the transaction	numeric	0
nameDest	customer who is the recipient of the transaction	nominal	0
oldbalanceDest	initial balance recipient before the transaction	numeric	0
newbalanceDest	new balance recipient after the transaction	numeric	0
isFraud	This is the transactions made by the fraudulent agents inside the simulation	nominal	0
isFlaggedFraud	The business model aims to control massive transfers from one account to another and flags illegal attempts	nominal	0

Table 1 - Features of dataset and missing values

Feature Names	Description	Average	Standard Deviation	Entropy	# of Values
step	maps a unit of time in the real world	243.397	142.332	5.5276	6.362.622
type	CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER	-	-	1.3077	6.362.622
amount	amount of the transaction in local currency	179861.90	603858.2	15.4085	6.362.622
nameOrig	customer who started the transaction	-	-	15.6639	6.362.622
oldbalanceOrig	initial balance before the transaction	833883.10	2888243	9.3923	6.362.622
newbalanceOrig	new balance after the transaction	855113.66	2924049	7.0845	6.362.622
nameDest	customer who is the recipient of the transaction	-	-	14.0318	6.362.622
oldbalanceDest	initial balance recipient before	1100701.66	3399180	9.3598	6.362.622
newbalanceDest	new balance recipient after the transaction	1224996.39	3674129	9.9373	6.362.622
isFraud	This is the transactions made by the fraudulent agents inside the simulation	-	-	0.0098	6.362.622
isFlaggedFraud	The business model aims to control massive transfers from one account to another and flags illegal attempts	-	-	0.000034	6.362.622

Table 2 - Features of dataset and descriptive values

Feature Names	Description	Type	Min. or Least Frequent	Max. or Most Frequent
step	maps a unit of time in the real world	numeric	1	743
type	CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER	nominal	DEBIT	CASH_OUT
amount	amount of the transaction in local currency	numeric	0	92.445.516,64
nameOrig	customer who started the transaction	nominal	C2066766136	C1999539787
oldbalanceOrg	initial balance before the transaction	numeric	0	59.585.040,37
newbalanceOrig	new balance after the transaction	numeric	0	49.585.040,37
nameDest	customer who is the recipient of the transaction	nominal	M917557255	C1286084959
oldbalanceDest	initial balance recipient before	numeric	0	356.015.889,35
newbalanceDest	new balance recipient after the transaction	numeric	0	356.179.278,92

isFraud	This is the transactions made by the fraudulent agents inside the simulation	nominal	0	1
isFlaggedFraud	The business model aims to control massive transfers from one account to another and flags illegal attempts	nominal	0	1

Table 3 - Features of dataset and descriptive values