GOBIERNO DE
EL SALVADOR

BCR

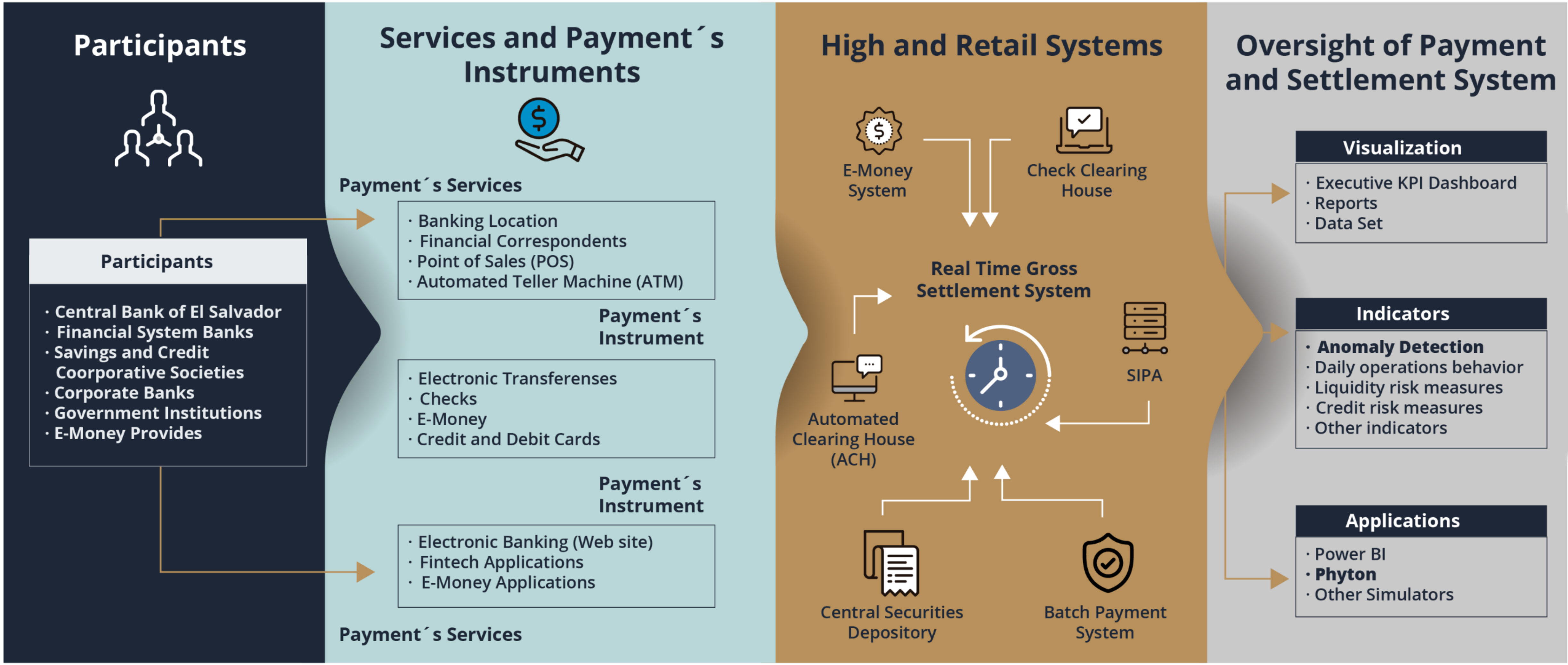
OVERSIGHT OF PAYMENT AND SETTLEMENT SYSTEM EL SALVADOR

Ing. Franklim Arevalo Guevara
Ing. William Medardo Rodríguez

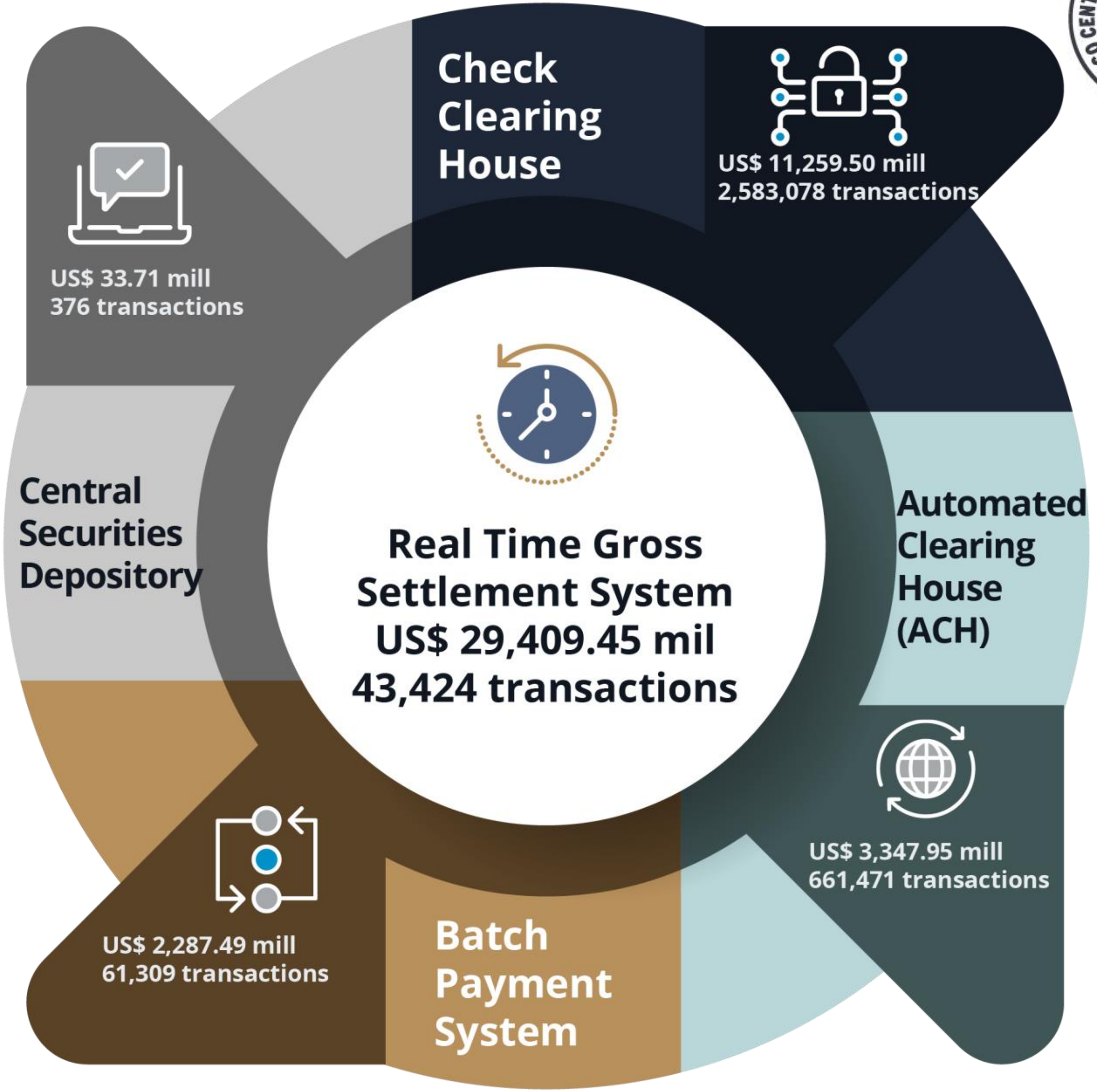
AGENDA

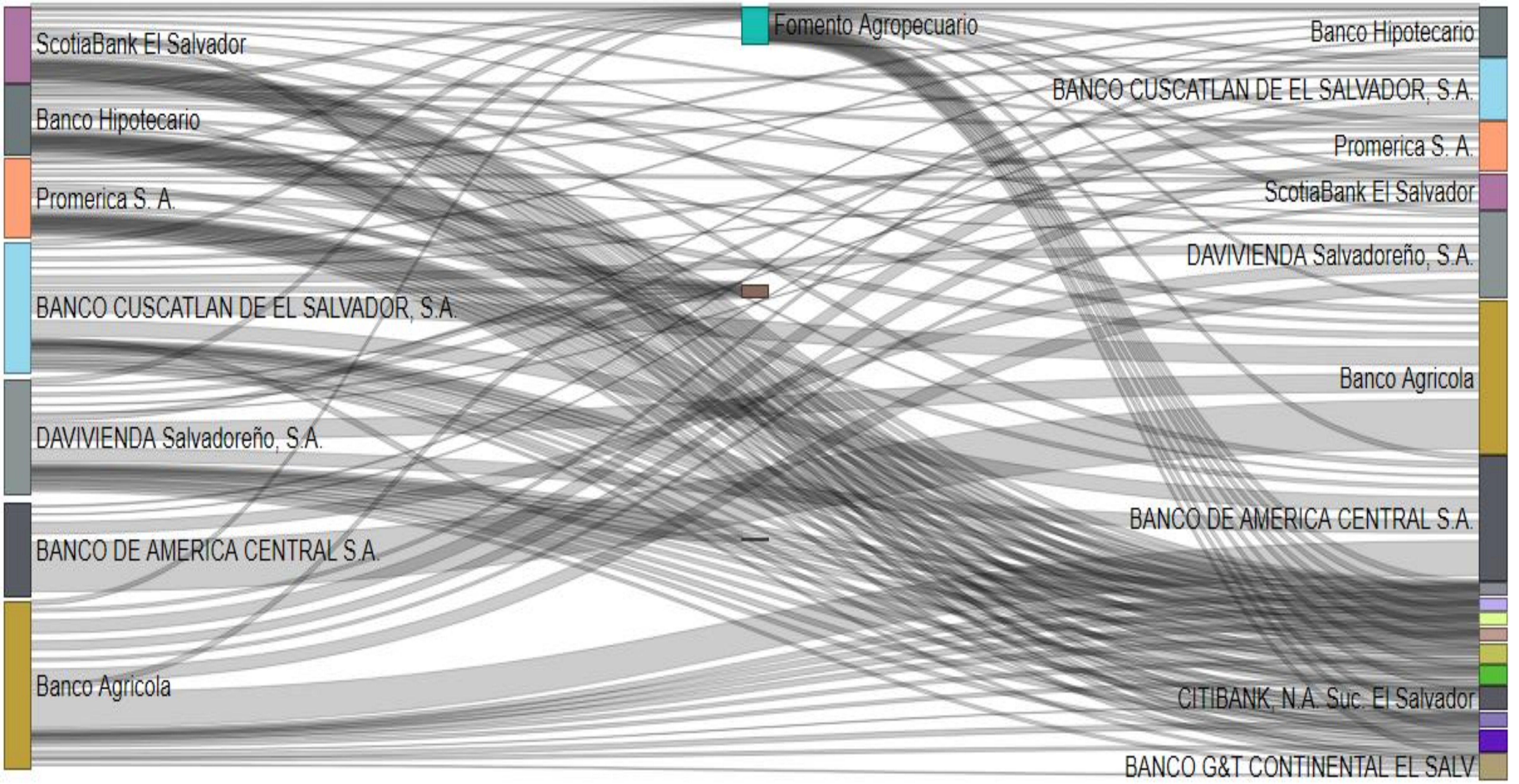
1. Oversight of Payment and Settlement Systems in El Salvador
2. Payment Systems in El Salvador
3. Objetive of the Project
4. Delvelopement of K-Means Method
5. Results of K-Means Method
6. Delvelopement of PCA Method
7. Results of PCA Method
8. Conclusions
9. Questions to discuss
10. References

1. Oversight of Payment and Settlement Systems



2. Payment transaction System in El Salvador




ScotiaBank El Salvador
Banco Hipotecario
Promerica S. A.
BANCO CUSCATLAN DE EL SALVADOR, S.A.
DAVIVIENDA Salvadoreño, S.A.
BANCO DE AMERICA CENTRAL S.A.
Banco Agricola
Fomento Agropecuario
Banco Hipotecario
BANCO CUSCATLAN DE EL SALVADOR, S.A.
Promerica S. A.
ScotiaBank El Salvador
DAVIVIENDA Salvadoreño, S.A.
Banco Agricola
BANCO DE AMERICA CENTRAL S.A.
CITIBANK, N.A. Suc. El Salvador
BANCO G&T CONTINENTAL EL SALV

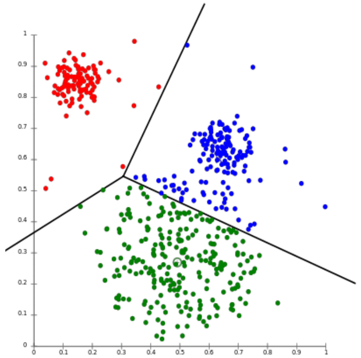
Oversight´s Pincipal Functions

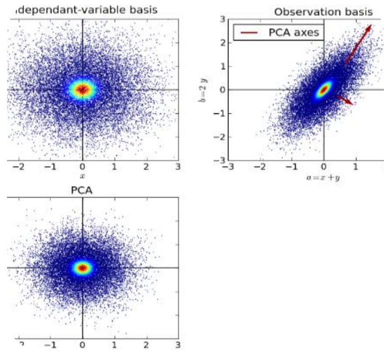
1. To analyze and monitor the operational behavior.
2. To develope and propose reforms to the regulatory framework.
3. To evaluate the principles for financial market infrastructures (IMF).
4. To control the Executive KPI Dashboard.
5. To develop reports on the performance of payment systems.
6. To coordinate the Payment System´s Modernization plan

3. Objective of the Project

Be able to apply Anomaly Detection on the Real Time Gross Settlement Systems to identify unusual payment behavior and help supervisors to initiate timely interventions.

What did we do?





Get Data Set from Real Time Gross Settlement System

Econometric Model

- Know the impact by "Type of transaction"

K-Means Method

- Anomaly Detection on the RTGS System (2 clusters)

PCA Method

- Anomaly Detection on the RTGS System (5 clusters)

Economic Model (Linear Regression)

GOBIERNO DE EL SALVADOR

BCR

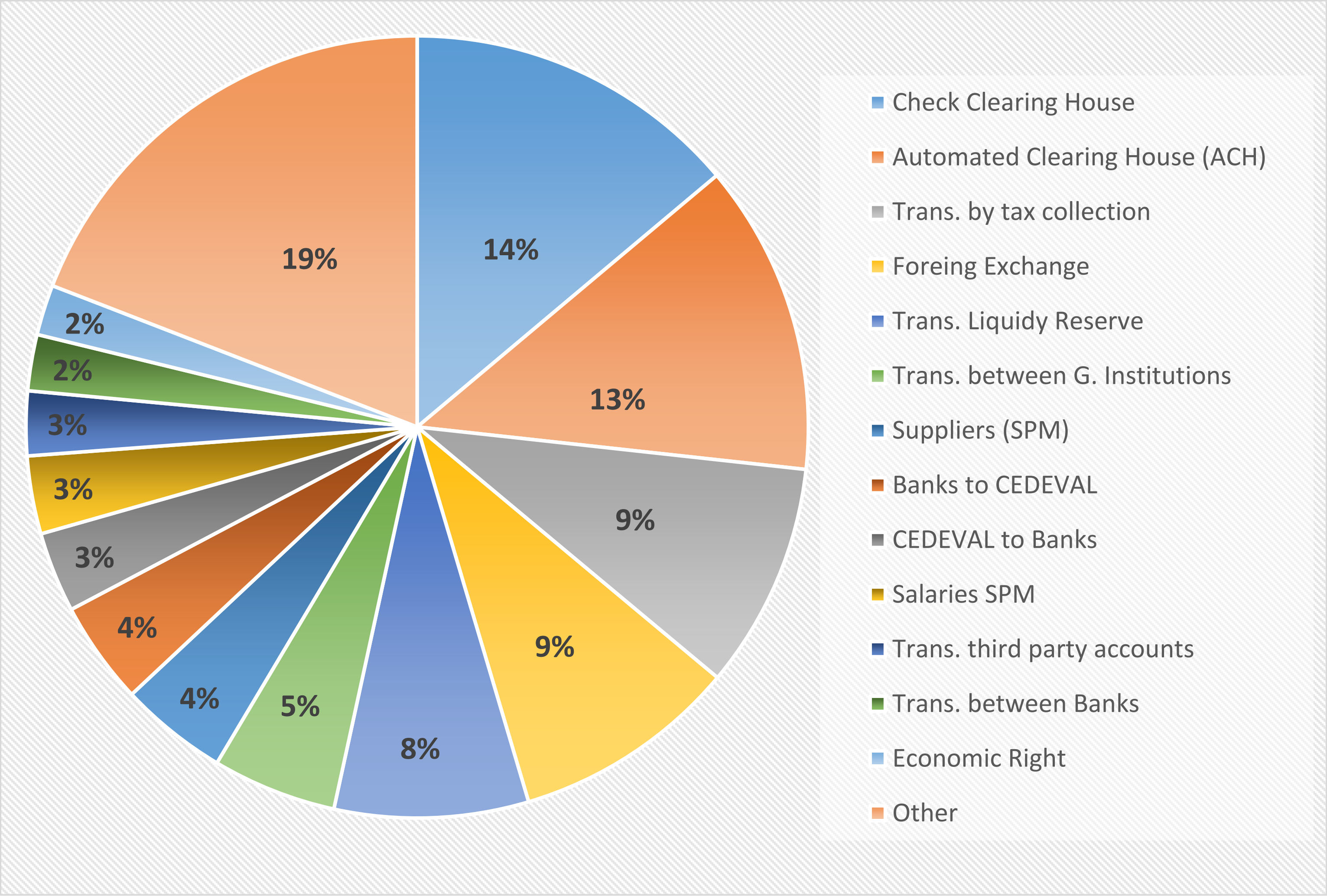
We decided to use a "Log-Log Linear Regression Model" becouse we wanted to stimate the coefficients of each "Type of transaction" to be able to know the impact of transaction made by "Type of transaction" in the Real-Time Gross Settlement System of El Salvador.

$$Log(LBTR) = \alpha 0 + \alpha 1\ Log(AVCENELIBOR) + \alpha 2\ Log(BDES) + \alpha 3\ Log(CICENELIBOR) + \alpha 4\ Log(CEDEVAL) + \alpha 5\ Log(DRRL) + \alpha 6\ Log(FILETES) + \alpha 7\ Log(IIITRL) + \alpha 8\ Log(IDIVISAS) + \alpha 9\ Log(LACH) + \alpha 10\ Log(LCCECH) + \alpha 11\ Log(PDEGOESINT) + \alpha 12\ Log(PGARANTIAS) + \alpha 13\ Log(PDGOPRINC) + \alpha 14\ Log(PLANILLASPM) + \alpha 15\ Log(TBBSAC) + \alpha 16\ Log(TCMBANCO) + \alpha 17\ Log(TRANSENTBANC) + \alpha 18\ Log(TRANSENTGOES) + \alpha 19\ Log(TRANSAFP) + \alpha 20\ Log(TRANSTERCER) + \alpha 21\ Log(AFPABANC) + \alpha 22\ Log(TRANSBREVES) + \alpha 23\ Log(RF) + \alpha 24\ DUMMIE + ui$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.045514	1.206682	2.523875	0.0234
LOG(ABONO_POR_VENC_CENELIBOR_...	-0.059216	0.043691	-1.355325	0.1954
LOG(BDES_A_TERC__1018_)	0.005874	0.004095	1.434635	0.1719
LOG(CARGO_POR_INVERSION_EN_CE...	0.095379	0.036816	2.590711	0.0205
LOG(CEDEVAL__1020_1091_1092_)	0.066786	0.018827	3.547352	0.0029
LOG(DEB_POR_RETIRO_DE_RL__2001_)	0.096437	0.041093	2.346819	0.0331
LOG(FDOS_POR_INV_EN_LETES__1039_)	-0.003936	0.006242	-0.630465	0.5379
LOG(III_TRAMO_DE_LA_RVA__DE_LIQU...	0.002839	0.016855	0.168415	0.8685
LOG(INGRESO_DE_DIVISAS__3001__)	0.135246	0.019834	6.818882	0.0000
LOG(LIQUIDACION_ACH__8001_)	0.090662	0.023819	3.806256	0.0017
LOG(LIQUIDACION_CCCECH__8004__)	0.148622	0.056219	2.643609	0.0184
LOG(PAGO_DE_DEUDA_GOES_INTERE...	-0.000465	0.011252	-0.041314	0.9676
LOG(PAGO_DE_GARANTIAS__1060_)	-0.005248	0.005408	-0.970455	0.3472
LOG(PAGO_DEUDA_GOES_PRINCIPAL_...	-0.000629	0.011039	-0.056939	0.9553
LOG(PAGO_PLANILLAS_SPM__1067_)	-0.001100	0.002738	-0.401806	0.6935
LOG(RF__1003_)	0.124952	0.029813	4.191191	0.0008
LOG(TRANS_B_BC_SAC__1007_)	0.055076	0.035284	1.560930	0.1394
LOG(TRANS_CUENTAS_MISMO_B__101...	0.004936	0.003692	1.336859	0.2012
LOG(TRANS_ENTRE_B__1026_)	0.035950	0.020056	1.792513	0.0932
LOG(TRANS_ENTRE_CUENTRAS_GOE...	0.123680	0.029726	4.160671	0.0008
LOG(TRANS_PARA_AFP__1008_)	0.045798	0.016316	2.806977	0.0133
LOG(TRANS_PARA_TERCE__1011_)	0.017541	0.017788	0.986135	0.3397
LOG(TRANSF_DE_AFPS_A_BANCOS__1...	0.009693	0.021469	0.451488	0.6581
LOG(TRANSF_DE_FONDOS_DE_BEVES...	0.000781	0.002894	0.269901	0.7909
DUMMIE	-0.022613	0.029876	-0.756907	0.4608

R-squared	0.987974	Mean dependent var	22.21984
Adjusted R-squared	0.968731	S.D. dependent var	0.133775
S.E. of regression	0.023655	Akaike info criterion	-4.381289
Sum squared resid	0.008394	Schwarz criterion	-3.325740
Log likelihood	112.6258	Hannan-Quinn criter.	-3.999636
F-statistic	51.34399	Durbin-Watson stat	1.754448
Prob(F-statistic)	0.000000		

Transaction made by the RTGS System



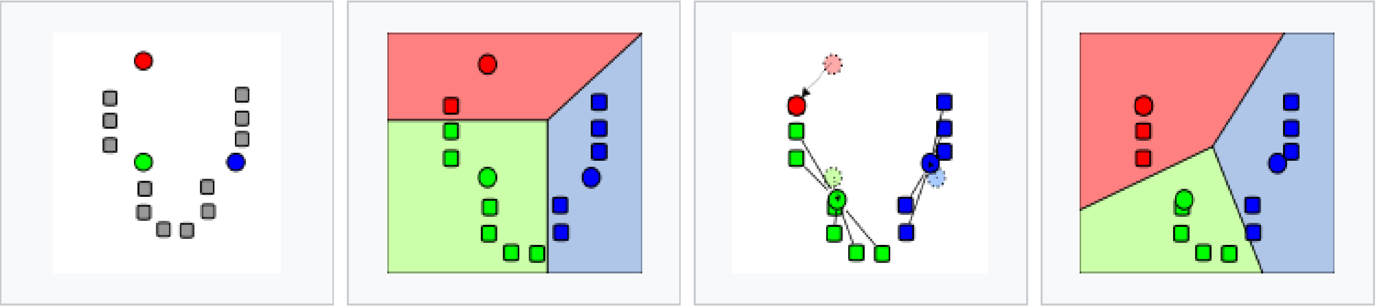
The Real Time Gross Settlement System has more than 114 different "Types of Operations", but just 12 of them (8.3%) represent more than 80% of the total amount.

In the first semester, the RTGS System made more than US$ 29,409.45 millions

4. K-means Method

- K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. the objective of K-means is simple: group similar data points together and discover underlying patterns.
- the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.
- k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. **Min** $\sum_{i=1}^{k} \left\|x_j - \mu i\right\|^2$

Demonstration of the standard algorithm



1. *k* initial "means" (in this case *k*=3) are randomly generated within the data domain (shown in color).

2. *k* clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.

3. The centroid of each of the *k* clusters becomes the new mean.

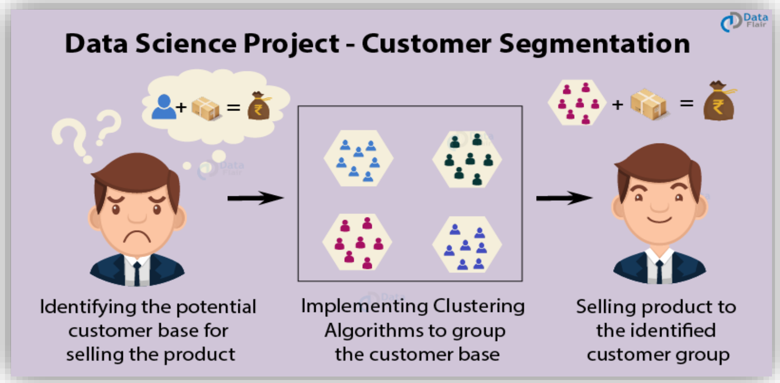
4. Steps 2 and 3 are repeated until convergence has been reached.

The K Means algorithm involves:

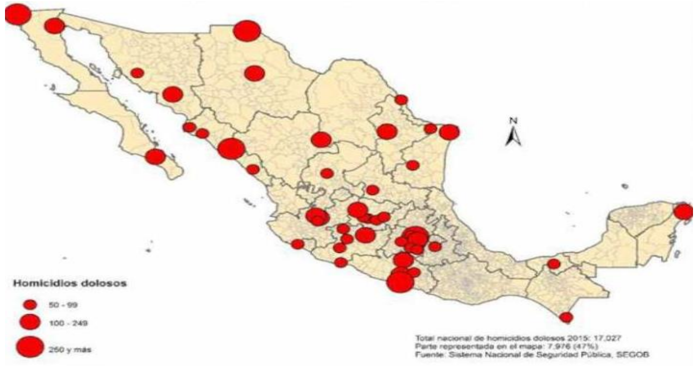
1. Choosing the number of clusters "k".
2. Randomly assign each point to a cluster.
3. Until clusters stop changing, repeat the following:
   - For each cluster, compute the cluster centroid by taking the mean vector of points in the cluster.
   - Assign each data point to the cluster for which the centroid is the closest.

Applications of K-Means Clustering

1. Customer segmentation.



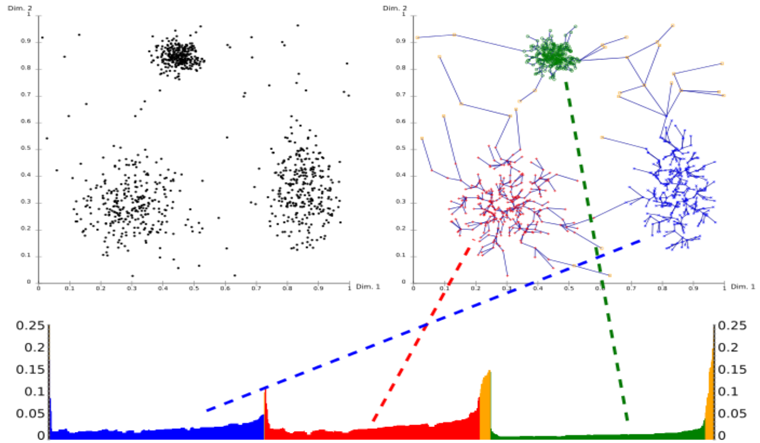
3. Identifying Crime Localities



2. Insurance Fraud Detection

Since insurance fraud can potentially have a multi-million dollar impact on a company, the ability to detect frauds is crucial.

4. Anomaly Detection



K-means Process

1: Import libraries

2: Load data set

3: Get and analyze statistical information

4: Sort instruction amount by operation type

5: Standardize the data set

6: Give to data set an specific format for K-means

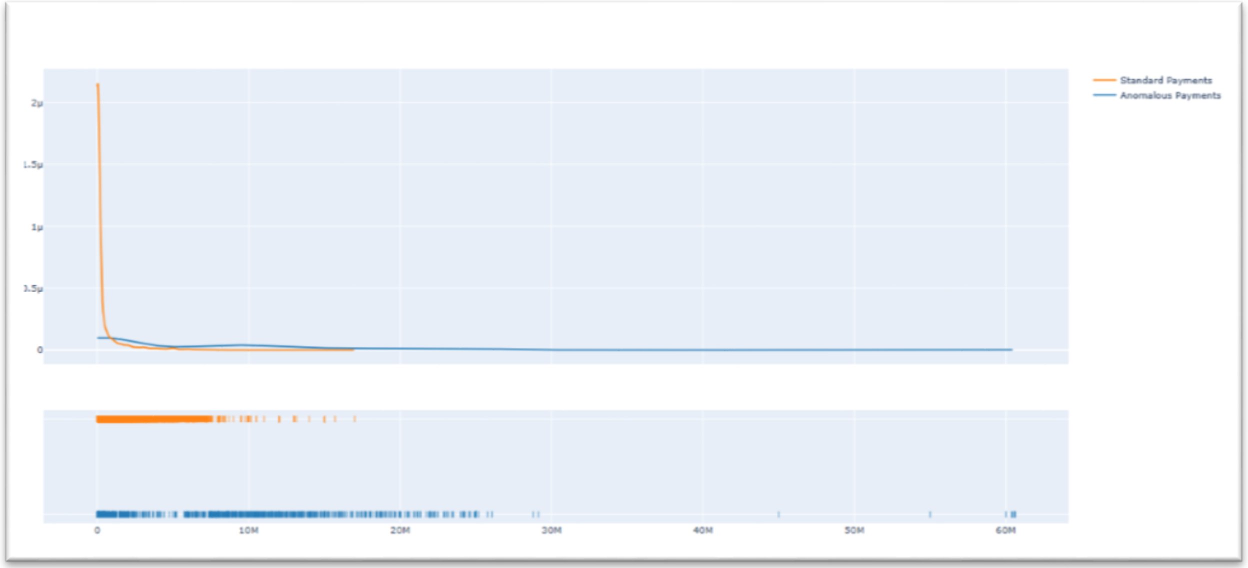
7: Feed and fit the model to train the K-means Method

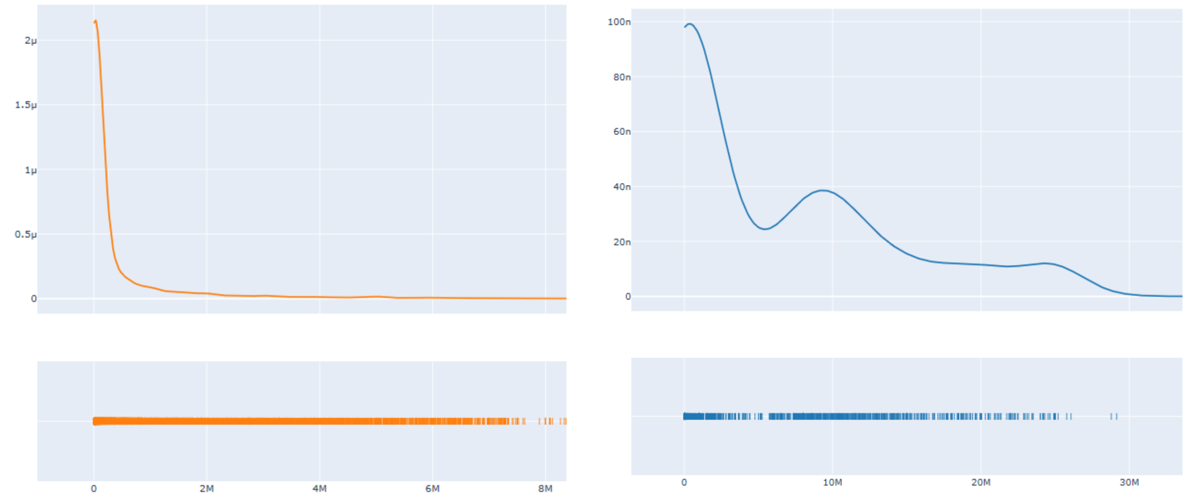
8: Show the results (Normal an suspicious anomalies payments)

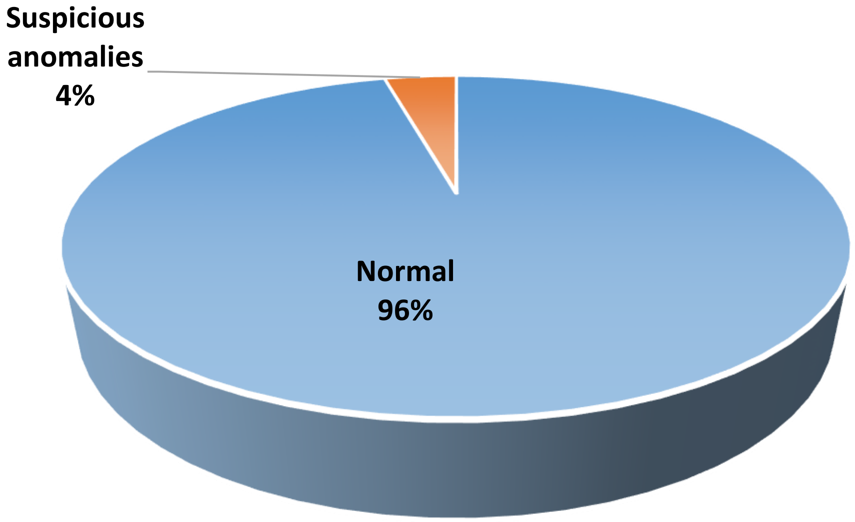
5. Results of K-Means Method

GOBIERNO DE EL SALVADOR

BCR



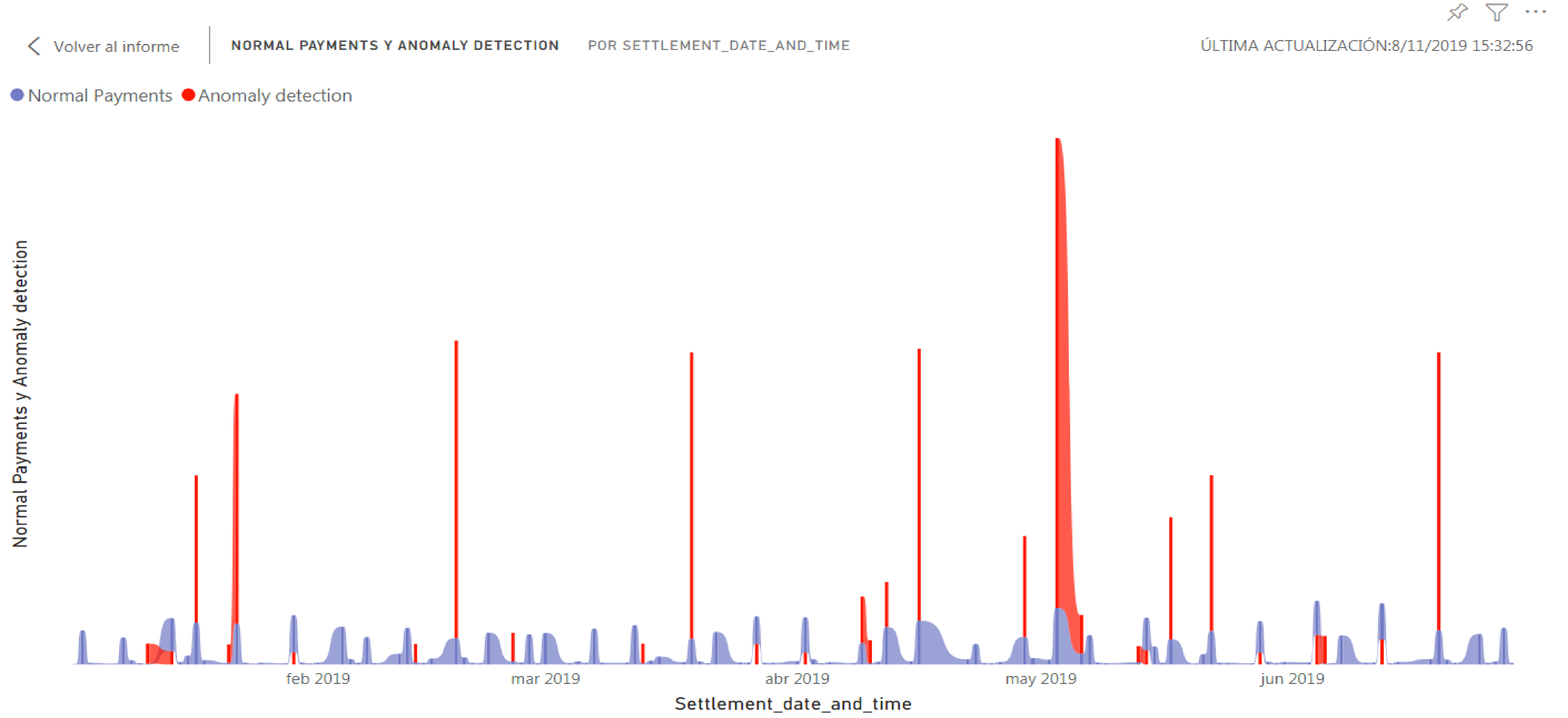




The results show us 2 clusters:

- Cluster 0: 96% of all the 100% data set have similarity payment´s features in the variable “amount”, related to Normal Payment´s transactions.
- Cluster 1: 4% of all the 100% data set have similarity payment´s features in the variable “amount”, related to suspicious anomalies payments

5. Results of K-Means Method (Power BI)



6. Principal Component Analysis Method

GOBIERNO DE EL SALVADOR

BCR

The Problem

- Data analysis requires to analyze multi dimensional data. We plot the data and find various patterns in it or use it to train some machine learning models.
- As the dimensions of data increases, the difficulty to visualize it and perform computations on it also increases.
- So, how to reduce the dimensions of a data-
  - Remove the redundant dimensions
  - Only keep the most important dimensions

What is PCA?

- Principal component analysis (PCA) is a technique used to emphasize variation and bring out strong patterns in a dataset. It's often used to make data easy to explore and visualize.
- PCA finds a new set of dimensions (or a set of basis of views) such that all the dimensions are orthogonal (and hence linearly independent) and ranked according to the variance of data along them.

	sepal length	sepal width	petal length	petal width
0	-0.900681	1.032057	-1.341272	-1.312977
1	-1.143017	-0.124958	-1.341272	-1.312977
2	-1.385353	0.337848	-1.398138	-1.312977
3	-1.506521	0.106445	-1.284407	-1.312977
4	-1.021849	1.263460	-1.341272	-1.312977

PCA
(2 components)

	principal component 1	princial component 2
0	-2.264542	0.505704
1	-2.086426	-0.655405
2	-2.367950	-0.318477
3	-2.304197	-0.575368
4	-2.388777	0.674767

6. Principal Component Analysis Method

GOBIERNO DE EL SALVADOR

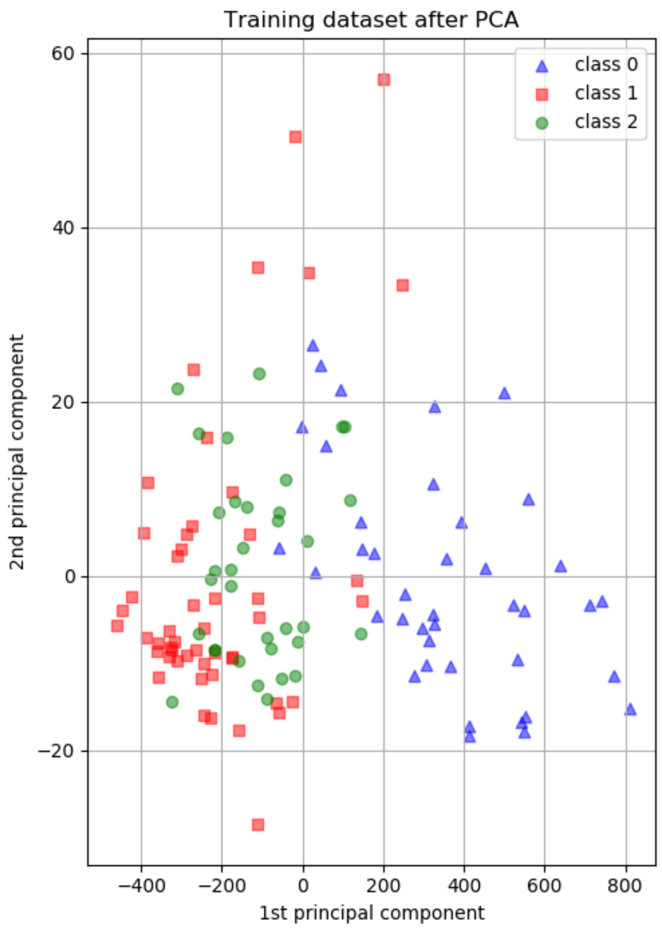
BCR

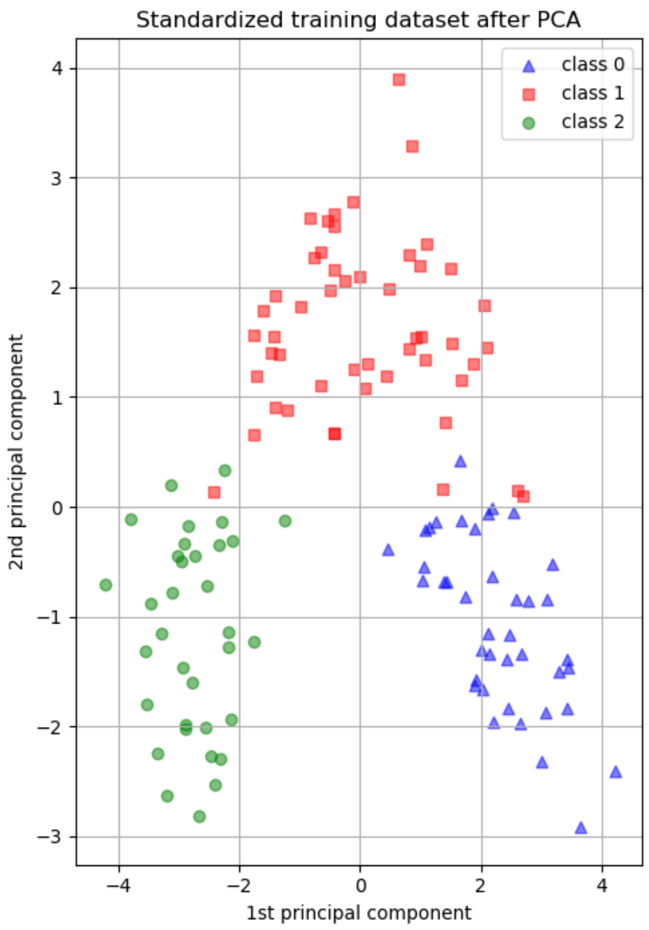
Applications
- PCA for Data Visualization
- PCA to Speed-up Machine Learning Algorithms

Explained Variance
- The explained variance tells you how much information (variance) can be attributed to each of the principal components.
- This is important as while you can convert 4 dimensional space to 2 dimensional space, you lose some of the variance (information) when you do this.

About data normalization

- Data need to be normalized before doing PCA because if we use data (features here) of different scales, we get misleading components.
- Use StandardScaler to help you standardize the dataset's features onto unit scale (mean = 0 and variance = 1) which is a requirement for the optimal performance of many machine learning algorithms.

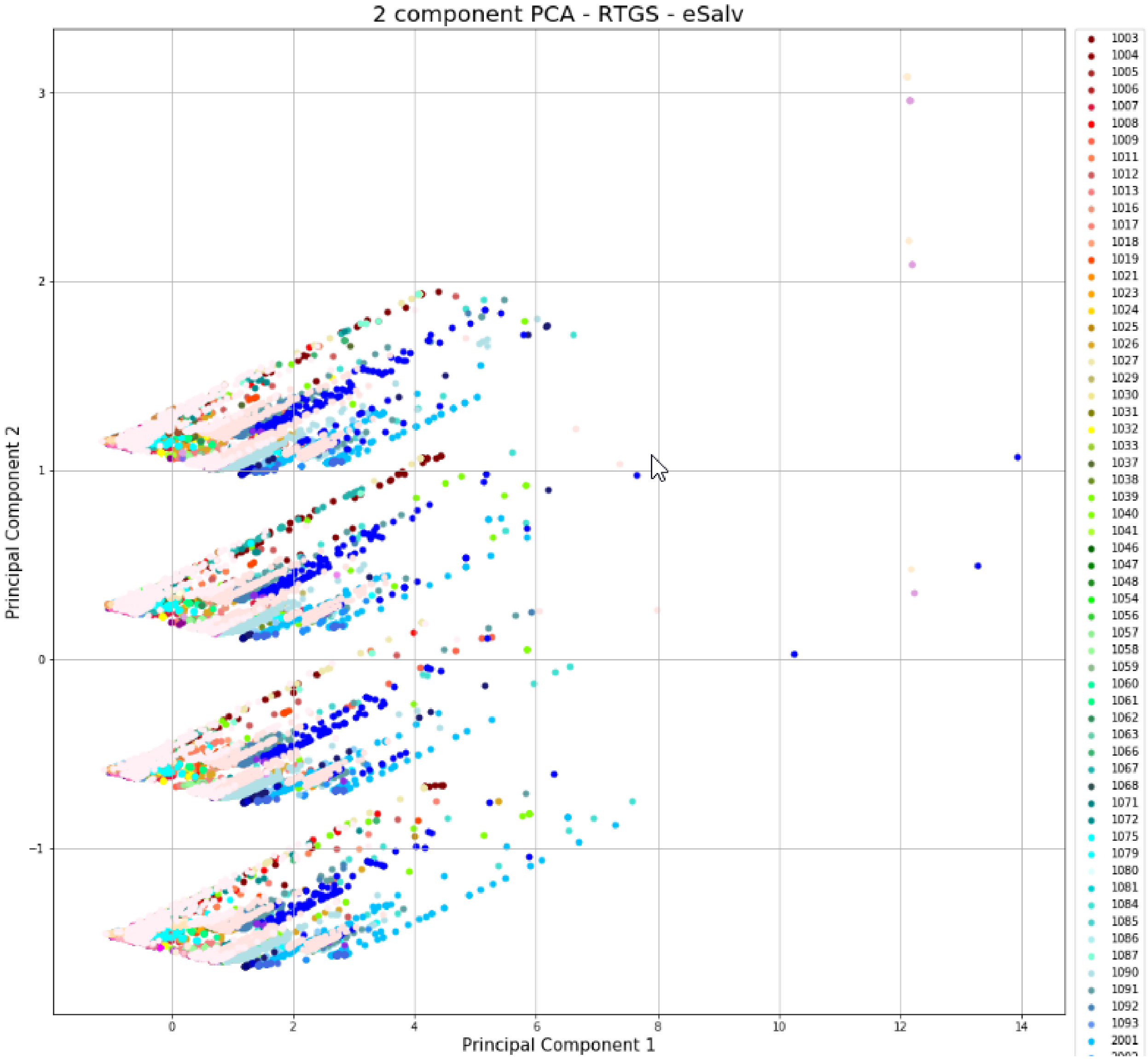


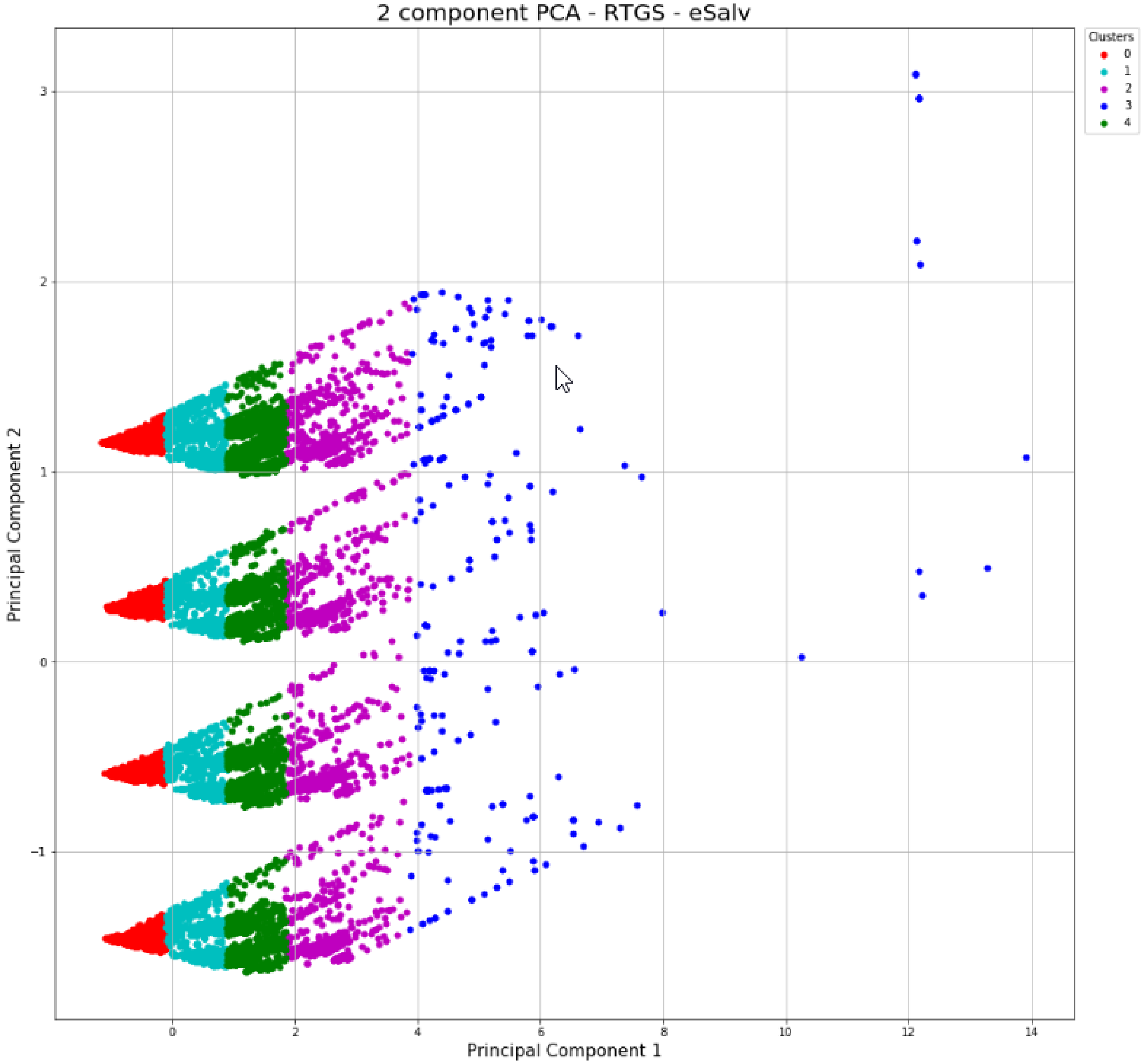


PCA + K-means (5 cluster) Process

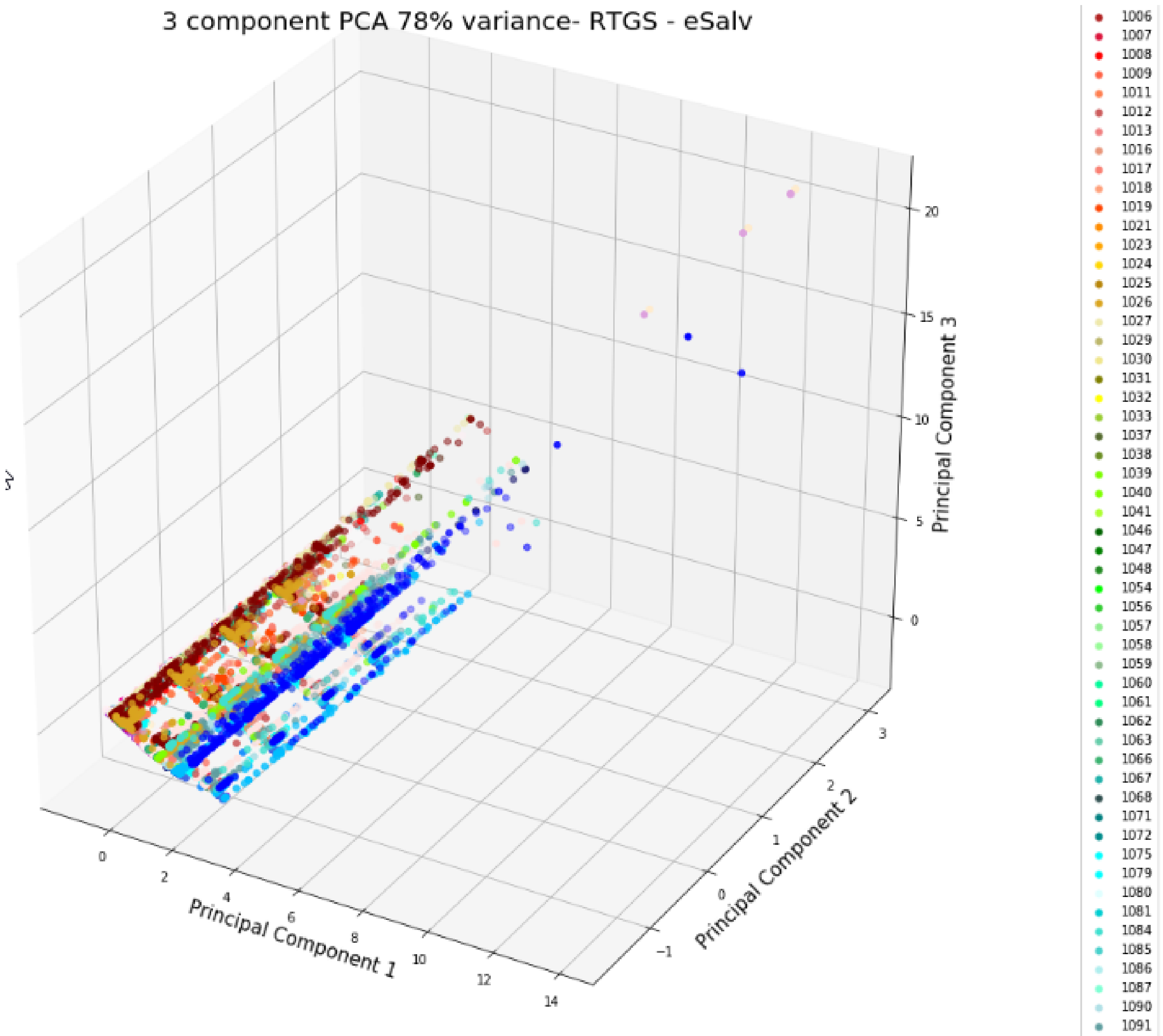
- 1: Import libraries
- 2: Load data set
- 4: Selecting key features
- 5: Standardizing the features
- 6: Defining features reduction (PCA components)
- 7: Applying PCA transform
- 8: Give to new components specific format for K-means
- 9: Define number of cluster for data visualization with PCA
- 10: Feed and fit the model to train the K-means Method
- 11: Analyze statistical information of the clusters
- 12: Show the results (Normal an suspicious anomalies payments)

7. Results of PCA + K-means (5 cluster) Method

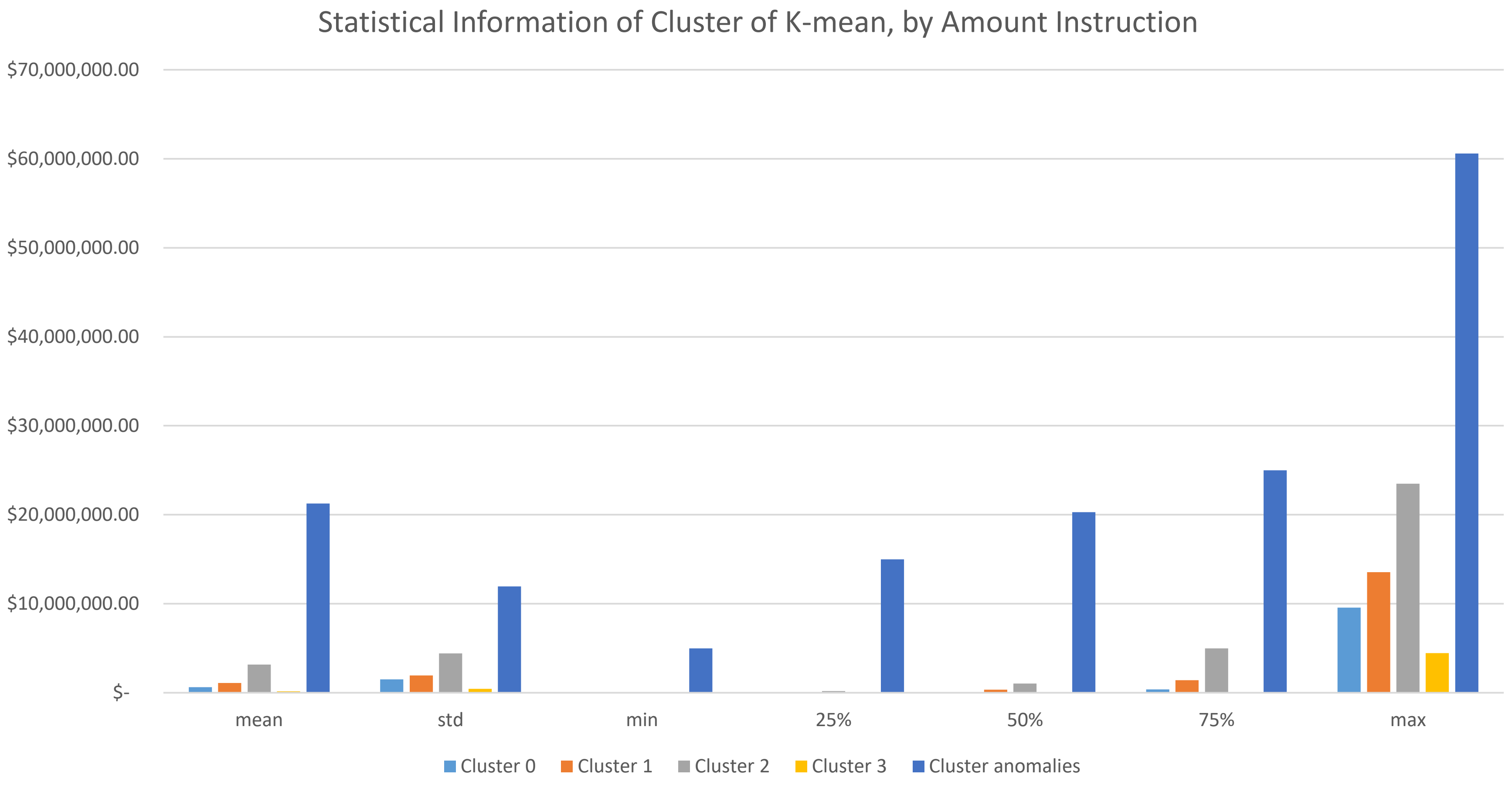


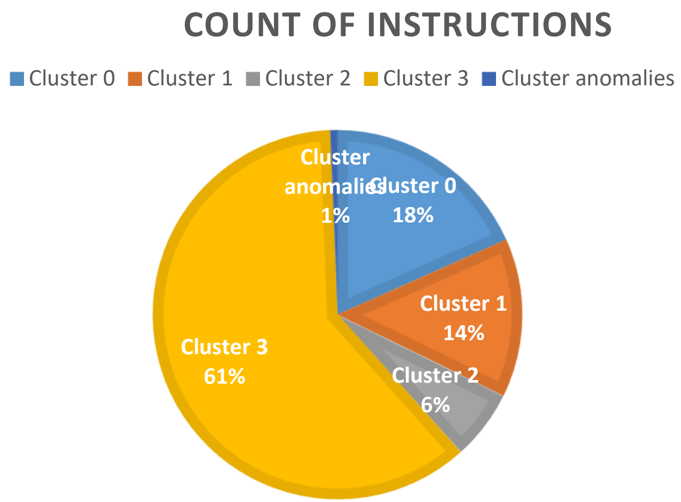


7. Results of PCA + K-means (5 cluster) Method

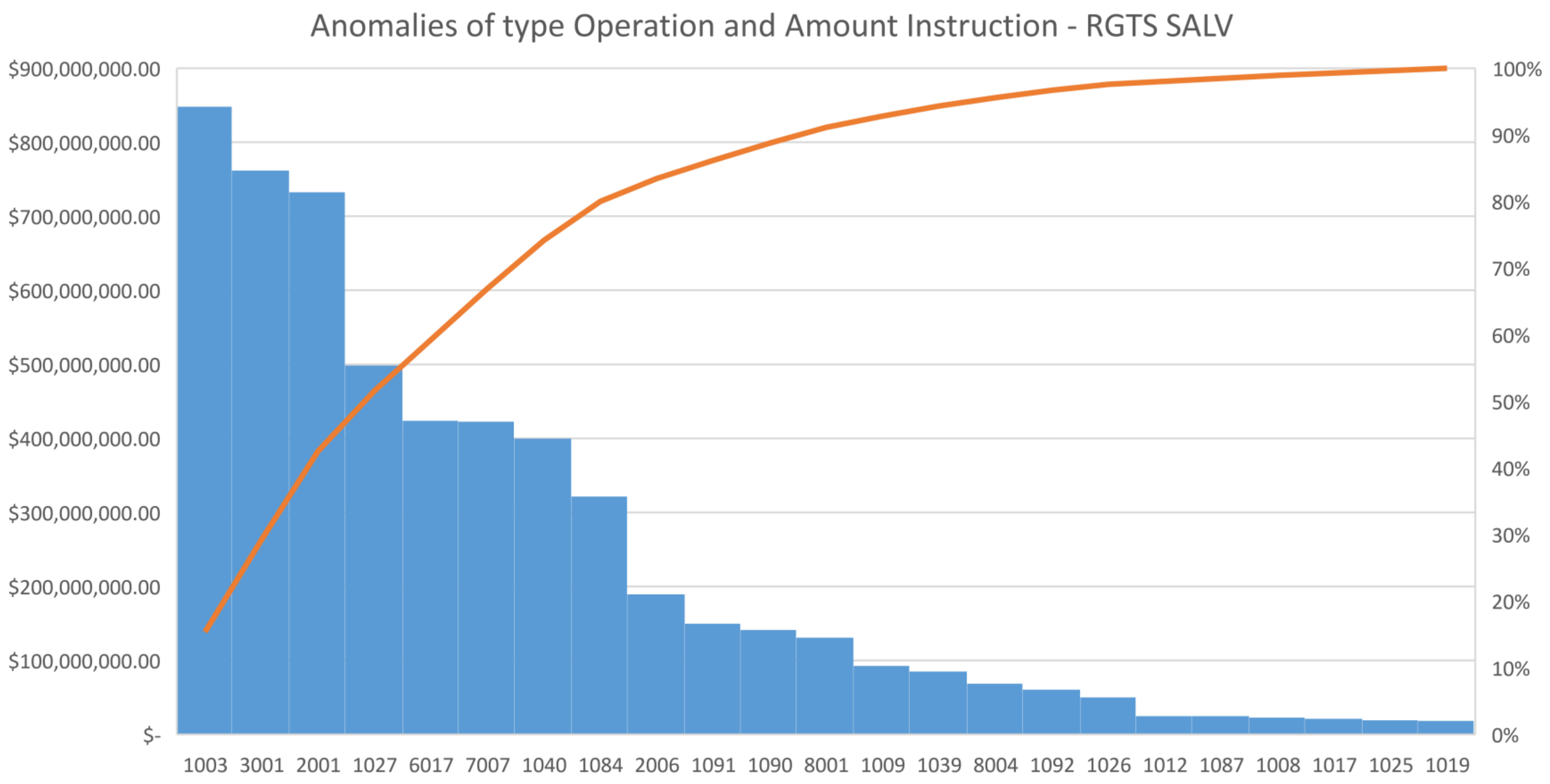


7. Results of PCA + K-means (5 cluster) Method





7. Results of PCA + K-means (5 cluster) Method



Type of Operations
1003 Transf. de Fdos. por Recaudación Fiscal
1008 Transf. de Fdos.de IFIs para AFPs
1009 Transf. de Fdos. de AFPs por Inver. en Certif. Previs. (FOP)
1012 Transf. de Fdos. entre cuentas del mismo Banco
1017 Transf. de Fdos. entre cuentas del BDES
1019 Transf. de Fdos. del BDES-FOP a Bancos para terceros
1025 Transf. de Fdos. de GOES a Bancos para Terceros
1026 Transf. de Fdos. entre Bancos
1027 Transf. de Fdos. entre cuentas del GOES
1039 Transferencia de Fdos por Inversión en LETES desmaterializados
1040 Transferencia de fdos por vencimiento de LETES desmaterializados
1084 Trans.de fondos CEDEVAL Derch Patrimoniales
1087 Transferencia de Fdos por Inversión en CETES desmaterializados
1090 Transf de Fdos de Instit a CEDEVAL
1091 Transf de Fdos Op de CEDEVAL para Instit
1092 Transf de Op de CEDEVAL para Terceros
2001 Debito por Retiro de Reserva de Liquidez
2006 Pago de Deuda GOES intereses
3001 Ingreso de divisas
6017 Abono por Venc CENELIBOR
7007 Cargo por Inversión en CENELIBOR
8001 Liquidación ACH
8004 Liquidacion Camara de Cheques

8. Conclusions

GOBIERNO DE EL SALVADOR

BCR

- In this exercise we used K-Means clustering technique to detect suspicious anomalies payments and we found some patterns on that could be anomalies payment´s transactions which can be helpful to detect new anomalies. We chose 6 months of payment´s transactions from the Real Time Gross Settlement System and then we refined the common properties among them. After gathering required information, we select 2 clusters according to insurance expert suggestion.
- The combined use of standardization techniques and PCA strengthen the quality and precision of the resulting analysis.
- The results of the first development indicates that the 96% of all the 100% data set are normal payments, and the 4% of all the 100% data set, we suspect that are anomalies payments.

8. Conclusions

- The use combined of k-means and PCA analysis using 5 number of cluster, indicates that the 1% of operations are unusual transactions.
- The second development shows that 80% of the anomalies, are found in 7 types of operations, related to unusual government transactions.
- The proposed methods succeed to extract some patterns and propositions which would be helpful to detect fraud anomalies payment in the future.

9. Questions to discuss

- How can we validate that the hypotheses has been performed are correct?
- How can we validate that the cluster number in k-means, or the key features for reductions in PCA, are suitable?
- Are we interpreting properly the graphs resulting from the model?

References

- Understanding Principal Component Analysis - Rishav Kumar
- https://medium.com/@aptrishu/understanding-principle-component-analysis-e32be0253ef0

- PCA using Python (scikit-learn) - Michael Galarnyk
https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60

- Principal Component Analysis - Victor Powell, Lewis Lehe
- http://setosa.io/ev/principal-component-analysis/

- Importance of Feature Scaling
- https://scikit-learn.org/stable/auto_examples/preprocessing/plot_scaling_importance.html#sphx-glr-auto-examples-preprocessing-plot-scaling-importance-py

GOBIERNO DE
EL SALVADOR

BCR

OVERSIGHT OF PAYMENT AND SETTLEMENT SYSTEM
EL SALVADOR

Ing. Franklim Arevalo Guevara
Ing. William Medardo Rodríguez