OVERSIGHT OF PAYMENT AND SETTLEMENT SYSTEM
EL SALVADOR
AGENDA

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

Participants
- Central Bank of El Salvador
- Financial System Banks
- Savings and Credit Coorporative Societies
- Corporate Banks
- Government Institutions
- E-Money Provides

Services and Payment´s Instruments

Payment´s Services
- Banking Location
- Financial Correspondents
- Point of Sales (POS)
- Automated Teller Machine (ATM)

Payment´s Instrument
- Electronic Transferenses
- Checks
- E-Money
- Credit and Debit Cards

Payment´s Services
- Electronic Banking (Web site)
- Fintech Applications
- E-Money Applications

High and Retail Systems

Real Time Gross Settlement System
- E-Money System
- Check Clearing House

Automated Clearing House (ACH)

Central Securities Depository

Batch Payment System

Oversight of Payment and Settlement System

Visualization
- Executive KPI Dashboard
- Reports
- Data Set

Indicators
- Anomaly Detection
- Daily operations behavior
- Liquidity risk measures
- Credit risk measures
- Other indicators

Applications
- Power BI
- Phyton
- Other Simulators
2. Payment transaction System in El Salvador
Oversight’s Principal Functions

1. To analyze and monitor the operational behavior.
2. To develop 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)
We decided to use a “Log-Log Linear Regression Model” because we wanted to estimate 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.

\[
\text{Log}(LBTR)= \alpha_0 + \alpha_1 \text{Log}(AVCENELIBOR) + \alpha_2 \text{Log}(BDES) + \alpha_3 \text{Log}(CICENELIBOR) + \alpha_4 \text{Log}(CEDEVAL) + \alpha_5 \text{Log}(DRRL) + \alpha_6 \text{Log}(FILETES) + \alpha_7 \text{Log}(IIITRL) + \alpha_8 \text{Log}(IDIVISAS) + \alpha_9 \text{Log}(LACH) + \alpha_{10} \text{Log}(LCCECH) + \alpha_{11} \text{Log}(PDEGOESINT) + \alpha_{12} \text{Log}(PGARANTIAS) + \alpha_{13} \text{Log}(PDGOPRINC) + \alpha_{14} \text{Log}(PLANILLASPM) + \alpha_{15} \text{Log}(TBBSAC) + \alpha_{16} \text{Log}(TCMBANCO) + \alpha_{17} \text{Log}(TRANSENTBANC) + \alpha_{18} \text{Log}(TRANSENTGOES) + \alpha_{19} \text{Log}(TRANSAFP) + \alpha_{20} \text{Log}(TRANSTEGR) + \alpha_{21} \text{Log}(AFPABANC) + \alpha_{22} \text{Log}(TRANSBREVES) + \alpha_{23} \text{Log}(RF) + \alpha_{24} \text{DUMMIE} + \epsilon
\]
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3.045514</td>
<td>1.206682</td>
<td>2.523875</td>
<td>0.0234</td>
</tr>
<tr>
<td>LOG(ABONO_POR_VENC_CENELIBOR__)</td>
<td>-0.059216</td>
<td>0.043691</td>
<td>-1.355325</td>
<td>0.1954</td>
</tr>
<tr>
<td>LOG(BDES_A_TERC__1018__)</td>
<td>0.005874</td>
<td>0.004095</td>
<td>1.434635</td>
<td>0.1719</td>
</tr>
<tr>
<td>LOG(CARGO_POR_INVERSION_EN_CENTRAL__)</td>
<td>0.095379</td>
<td>0.036816</td>
<td>2.590711</td>
<td>0.0205</td>
</tr>
<tr>
<td>LOG(CEDEVAL__1020_1091_1092__)</td>
<td>0.066786</td>
<td>0.018827</td>
<td>3.547352</td>
<td>0.0029</td>
</tr>
<tr>
<td>LOG(DEB_POR_RETIRO_DE_RL__2001__)</td>
<td>0.096437</td>
<td>0.041093</td>
<td>2.346819</td>
<td>0.0331</td>
</tr>
<tr>
<td>LOG(FDOS_POR_INV_EN_LETES__1039__)</td>
<td>-0.003936</td>
<td>0.006242</td>
<td>-0.630465</td>
<td>0.5379</td>
</tr>
<tr>
<td>LOG(III_TRAMO_DE_LA_RVA_DE_LIQUIDACION__)</td>
<td>0.002839</td>
<td>0.016855</td>
<td>0.168415</td>
<td>0.8685</td>
</tr>
<tr>
<td>LOG(INGRESO_DE_DIVISAS__3001__)</td>
<td>0.135246</td>
<td>0.019834</td>
<td>6.818882</td>
<td>0.0000</td>
</tr>
<tr>
<td>LOG(LIQUIDACION_ACH__8001__)</td>
<td>0.090662</td>
<td>0.023819</td>
<td>3.806256</td>
<td>0.0017</td>
</tr>
<tr>
<td>LOG(LIQUIDACION_CCCECH__8004__)</td>
<td>0.148622</td>
<td>0.056219</td>
<td>2.643609</td>
<td>0.0184</td>
</tr>
<tr>
<td>LOG(PAGO_DE_DEUDA_GOES_INTERE__1060__)</td>
<td>-0.005248</td>
<td>0.005408</td>
<td>-0.970455</td>
<td>0.3472</td>
</tr>
<tr>
<td>LOG(PAGO_DE_GARANTIAS__1060__)</td>
<td>-0.005248</td>
<td>0.005408</td>
<td>-0.970455</td>
<td>0.3472</td>
</tr>
<tr>
<td>LOG(PAGO_DE_GARANTIAS__1060__)</td>
<td>-0.005248</td>
<td>0.005408</td>
<td>-0.970455</td>
<td>0.3472</td>
</tr>
<tr>
<td>LOG(PAGO_PLANILLAS_SPM__1067__)</td>
<td>-0.001100</td>
<td>0.002738</td>
<td>-0.401806</td>
<td>0.6935</td>
</tr>
<tr>
<td>LOG(RF__1003__)</td>
<td>0.124952</td>
<td>0.029813</td>
<td>4.191191</td>
<td>0.0008</td>
</tr>
<tr>
<td>LOG(TRANS_B_BC_SAC__1007__)</td>
<td>0.055076</td>
<td>0.035284</td>
<td>1.560930</td>
<td>0.1394</td>
</tr>
<tr>
<td>LOG(TRANS_CUENTAS_MISMO_B__1011__)</td>
<td>0.004936</td>
<td>0.003692</td>
<td>1.336859</td>
<td>0.2012</td>
</tr>
<tr>
<td>LOG(TRANS_ENTRE_B__1026__)</td>
<td>0.035950</td>
<td>0.020056</td>
<td>1.792513</td>
<td>0.0932</td>
</tr>
<tr>
<td>LOG(TRANS_ENTRE_CUENTRAS_GOE__1020__)</td>
<td>0.123680</td>
<td>0.029726</td>
<td>4.160671</td>
<td>0.0008</td>
</tr>
<tr>
<td>LOG(TRANS_PARA_AFP__1008__)</td>
<td>0.045798</td>
<td>0.016316</td>
<td>2.806977</td>
<td>0.0133</td>
</tr>
<tr>
<td>LOG(TRANS_PARA_TEC__1011__)</td>
<td>0.017541</td>
<td>0.017788</td>
<td>0.986135</td>
<td>0.3397</td>
</tr>
<tr>
<td>LOG(TRANSF_DE_AFPS_A_BANCOS__1001__)</td>
<td>0.009693</td>
<td>0.021469</td>
<td>0.451488</td>
<td>0.6581</td>
</tr>
<tr>
<td>LOG(TRANSF_DE_FONDOS_DE_BEVES__)</td>
<td>0.000781</td>
<td>0.002894</td>
<td>0.269907</td>
<td>0.7909</td>
</tr>
<tr>
<td>DUMMY</td>
<td>-0.022613</td>
<td>0.029876</td>
<td>-0.756907</td>
<td>0.4608</td>
</tr>
</tbody>
</table>

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
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 $\sum_{i=1}^{k} \|x_j - \mu_i\|^2$

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


Data Science Project - Customer Segmentation

- Identifying the potential customer base for selling the product
- Implementing Clustering Algorithms to group the customer base
- Selling product to the identified customer group

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.

3. Identifying Crime Localities

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

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

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.
6. Principal Component Analysis Method

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
8: Give to new components specific format for K-means
7: Applying PCA transform
6: Defining features reduction (PCA components)
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)

20
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

**Statistical Information of Cluster of K-mean, by Amount Instruction**

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Percent</th>
<th>COUNT OF INSTRUCTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 0</td>
<td>18%</td>
<td>Cluster 0 18%</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>14%</td>
<td>Cluster 1 14%</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>6%</td>
<td>Cluster 2 6%</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>61%</td>
<td>Cluster 3 61%</td>
</tr>
<tr>
<td>Cluster anomalies</td>
<td>1%</td>
<td>Cluster anomalies 1%</td>
</tr>
</tbody>
</table>

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

- Mean:
  - Cluster 0: $10,000,000.00
  - Cluster 1: $20,000,000.00
  - Cluster 2: $30,000,000.00
  - Cluster 3: $40,000,000.00
  - Cluster anomalies: $50,000,000.00

- Std:
  - Cluster 0: $1,000,000.00
  - Cluster 1: $2,000,000.00
  - Cluster 2: $3,000,000.00
  - Cluster 3: $4,000,000.00
  - Cluster anomalies: $5,000,000.00

- Min:
  - Cluster 0: $5,000,000.00
  - Cluster 1: $10,000,000.00
  - Cluster 2: $15,000,000.00
  - Cluster 3: $20,000,000.00
  - Cluster anomalies: $25,000,000.00

- 25%:
  - Cluster 0: $7,500,000.00
  - Cluster 1: $12,500,000.00
  - Cluster 2: $17,500,000.00
  - Cluster 3: $22,500,000.00
  - Cluster anomalies: $27,500,000.00

- 50%:
  - Cluster 0: $10,000,000.00
  - Cluster 1: $15,000,000.00
  - Cluster 2: $20,000,000.00
  - Cluster 3: $25,000,000.00
  - Cluster anomalies: $30,000,000.00

- 75%:
  - Cluster 0: $12,500,000.00
  - Cluster 1: $17,500,000.00
  - Cluster 2: $22,500,000.00
  - Cluster 3: $27,500,000.00
  - Cluster anomalies: $32,500,000.00

- Max:
  - Cluster 0: $50,000,000.00
  - Cluster 1: $60,000,000.00
  - Cluster 2: $70,000,000.00
  - Cluster 3: $80,000,000.00
  - Cluster anomalies: $90,000,000.00
### 7. Results of PCA + K-means (5 cluster) Method

<table>
<thead>
<tr>
<th>Type of Operations</th>
<th>Amount Instruction - RGTS SALV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1003  Transf. de Fdos. por Recaudación Fiscal</td>
<td>$900,000,000.00</td>
</tr>
<tr>
<td>1008  Transf. de Fdos.de IFIs para AFPs</td>
<td>$800,000,000.00</td>
</tr>
<tr>
<td>1009  Transf. de Fdos. de AFPs por Inver. en Certif. Previs. (FOP)</td>
<td>$700,000,000.00</td>
</tr>
<tr>
<td>1012  Transf. de Fdos. entre cuentas del mismo Banco</td>
<td>$600,000,000.00</td>
</tr>
<tr>
<td>1017  Transf. de Fdos. entre cuentas del BOES</td>
<td>$500,000,000.00</td>
</tr>
<tr>
<td>1019  Transf. de Fdos. del BOES-FOP a Bancos para terceros</td>
<td>$400,000,000.00</td>
</tr>
<tr>
<td>1025  Transf. de Fdos. de GOES a Bancos para Terceros</td>
<td>$300,000,000.00</td>
</tr>
<tr>
<td>1026  Transf. de Fdos. entre Bancos</td>
<td>$200,000,000.00</td>
</tr>
<tr>
<td>1027  Transf. de Fdos. entre cuentas del GOES</td>
<td>$100,000,000.00</td>
</tr>
<tr>
<td>1039  Transferencia de Fdos por Inversión en LETES desmaterializados</td>
<td>$100,000.00</td>
</tr>
<tr>
<td>1040  Transferencia de fdos por vencimiento de LETES desmaterializados</td>
<td></td>
</tr>
<tr>
<td>1084  Trans.de fondos CEDEVAL Derch Patrimoniales</td>
<td></td>
</tr>
<tr>
<td>1087  Transferencia de Fdos por Inversión en CETES desmaterializados</td>
<td></td>
</tr>
<tr>
<td>1090  Transf de Fdos de Instit a CEDEVAL</td>
<td></td>
</tr>
<tr>
<td>1091  Transf de Fdos Op de CEDEVAL para Inst</td>
<td></td>
</tr>
<tr>
<td>1092  Transf de Op de CEDEVAL para Terceros</td>
<td></td>
</tr>
<tr>
<td>2001  Debito por Retiro de Reserva de Liquidez</td>
<td></td>
</tr>
<tr>
<td>2006  Pago de Deuda GOES intereses</td>
<td></td>
</tr>
<tr>
<td>3001  Ingreso de divisas</td>
<td></td>
</tr>
<tr>
<td>6017  Abono por Venc CENELIBOR</td>
<td></td>
</tr>
<tr>
<td>7007  Cargo por Inversión en CENELIBOR</td>
<td></td>
</tr>
<tr>
<td>8001  Liquidación ACH</td>
<td></td>
</tr>
<tr>
<td>8004  Liquidacion Camara de Cheques</td>
<td></td>
</tr>
</tbody>
</table>

Anomalies of type Operation and Amount Instruction - RGTS SALV
8. Conclusions

• 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
OVERSIGHT OF PAYMENT AND SETTLEMENT SYSTEM
EL SALVADOR

Ing. Franklim Arevalo Guevara
Ing. William Medardo Rodríguez