

*The Application of Exploratory Data Analysis  
(EDA) in Auditing*

Qi Liu

Ph.D. Candidate (A.B.D.)

Dept. of Accounting & Information Systems

Rutgers University

**28<sup>th</sup> WCARS**

November 9, 2013

# Outline

- ❖ Introduction
- ❖ An overview of EDA concept
- ❖ EDA in Auditing
- ❖ An application of EDA in auditing – A credit card retention case
- ❖ Future Research

# Introduction

## ❖ Motivation

- Audit is a data intensive process; data analysis plays an important role in audit process.
- Current data analysis approaches used in auditing process focus on validating predefined audit objectives, which can not discover unaware risks from the data.
- EDA is often linked to detective work, and one of its objectives is to identify outliers.
- Even though some EDA techniques have been used in some auditing procedures, EDA has never been systematically employed in auditing.

## ❖ Contribution

- This research contributes to the auditing literature by taking the first cut to use exploratory data analysis in auditing and illustrate a real-world application in audit process.

# Definition of EDA

- ❖ Exploratory data analysis (EDA) is a data analysis approach emphasizing on pattern recognition and hypothesis generation.



# EDA vs CDA

- ❖ Confirmatory Data Analysis (CDA) is a widely used data analysis approach emphasizing on experimental design, significance testing, estimation, and prediction (Good, 1983).

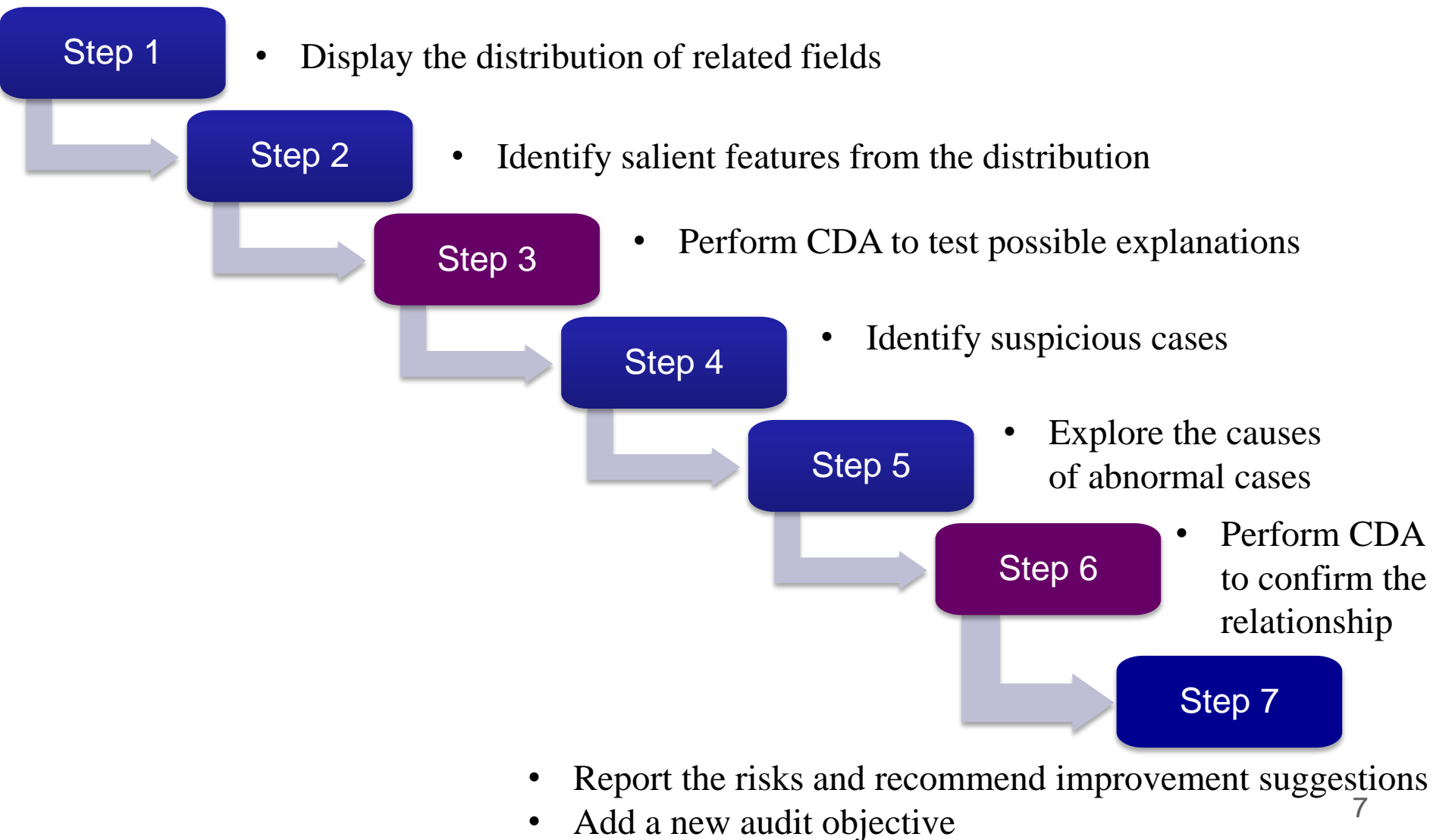
	Exploratory Data Analysis (EDA)	Confirmatory Data Analysis (CDA)
<b>Reasoning Type</b>	Inductive	Deductive
<b>Goal</b>	Pattern Recognition and Hypothesis generation	Estimation, Modeling, Hypothesis testing
<b>Applied Data</b>	Observation Data (data collected without well-defined hypothesis)	Experimental data (data collected through formally designed experiments)
<b>Techniques</b>	Descriptive Statistics, Data Visualization, Clustering Analysis, Process Mining...	Traditional statistical techniques of inference, significance, and confidence
<b>Advantages</b>	<ul style="list-style-type: none"> <li>• No assumptions required</li> <li>• Promotes deeper understanding of the data</li> </ul>	<ul style="list-style-type: none"> <li>• Precise</li> <li>• Well-established theory and methods</li> </ul>
<b>Disadvantages</b>	<ul style="list-style-type: none"> <li>• No conclusive answers</li> <li>• Difficult to avoid bias produced by overfitting</li> </ul>	<ul style="list-style-type: none"> <li>• Required unrealistic assumptions</li> <li>• Difficult to notice unexpected results</li> </ul>

## Current Applications of EDA

- ❖ Since 1980s, EDA has been applied to diversified disciplines such as interior design, marketing, industrial engineering, and geography (*Chen et al., 2011; Nayaka and Yano, 2010; Koschat and Sabavala, 1994; Wesley et al., 2006; De Mast and Trip, 2007, 2009*).
- ❖ A framework to apply EDA in practical problem solving issues include: (1) display the data; (2) identify salient features; (3) interpret salient features (De Mast and Kemper, 2009).



# Framework to apply EDA in auditing



# Credit Card Retention Case

## ❖ Purpose

- Demonstrate the benefits of applying EDA in audit process
- Provide a real example to support the proposed guidelines

❖ **Scenario:** Clients call the bank asking for a reduction of their card fees. Bank representatives offer discounts to clients to retain their accounts.

❖ **Objectives:** identify the situations of loss of revenue in the negotiation of fees caused by bank representatives, as b:

- bank representatives offer higher discounts than allowed
- bank representatives usually offer the highest allowable discounts without putting enough efforts to negotiate lower discounts
- bank representatives offer discounts without any negotiation with the clients



# Data Description

## Data (Retention Dataset)

- Each record represents a customer call
- 195,694 records
- 162 fields
- Time frame: January, 2012

## Selected Attributes

- Original fee (VLR\_ANUIDADE\_G)
- Actual fee (\_Valor da Anuidade de Saída)
- Agent identification (Funcional do Agente)
- Supervisor identification (Funcional do Supervisor)
- Location of the customer service center (Polo de Atendimento)
- Call duration (Tempo de Atendimento de Retenti)

# Methodology

## ❖ Data Preprocess

- Discount Calculation

$$\text{Discount} = \frac{(\text{Original fee} - \text{Actual fee})}{\text{Original fee}} \times 100\%$$

## ❖ Applied EDA techniques

- Descriptive Statistics
- Data Visualization
- Data Transformation

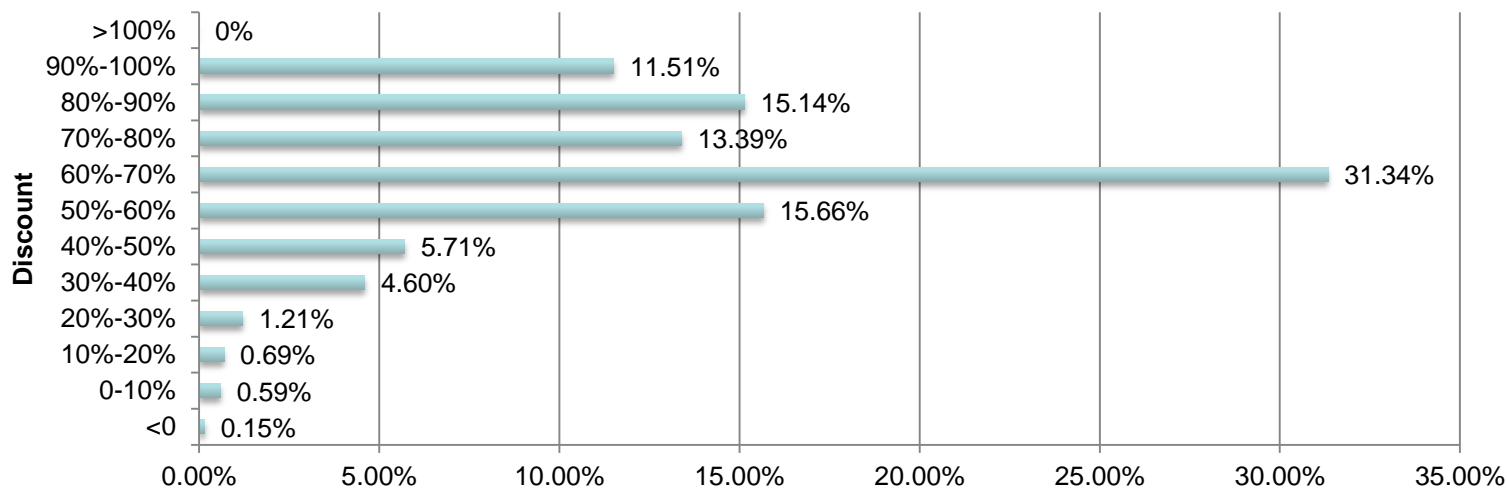
# Results Analysis (1/8)

## ❖ Policy-violating bank representatives and negative discounts

- Bank policy allows bank representatives to offer discounts up to 100% of the annuity to retain the customer

Field Name	Mean	Median	Minimum	Maximum	Standard deviation
Discount	-2.326.04	60	-27,944,522.22	100.00	219933.88

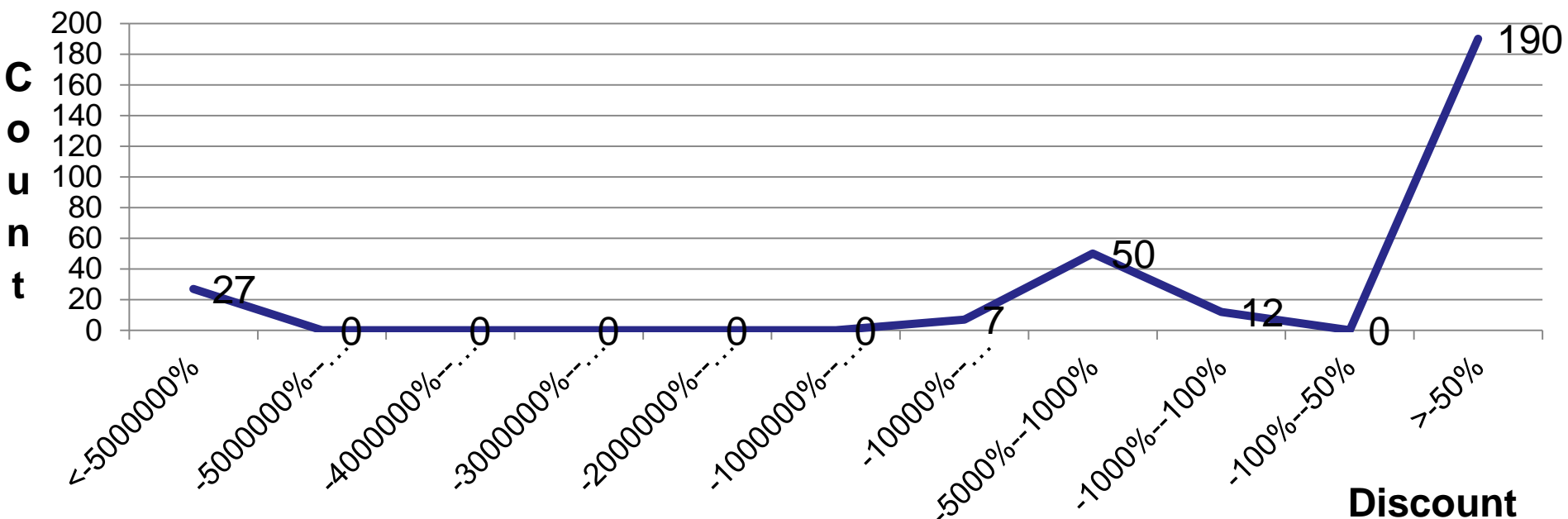
*Descriptive statistics of discounts*



*Frequency distribution of discounts*

## Results Analysis (2/8)

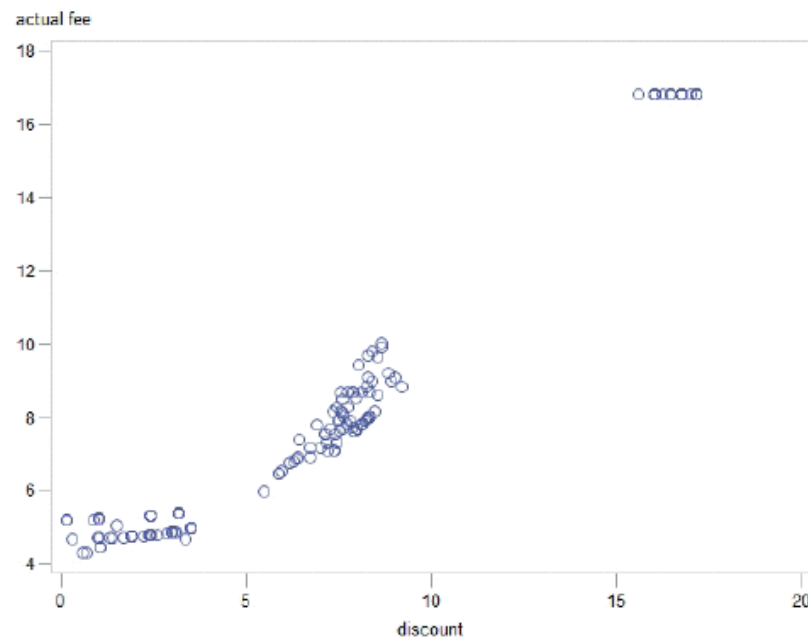
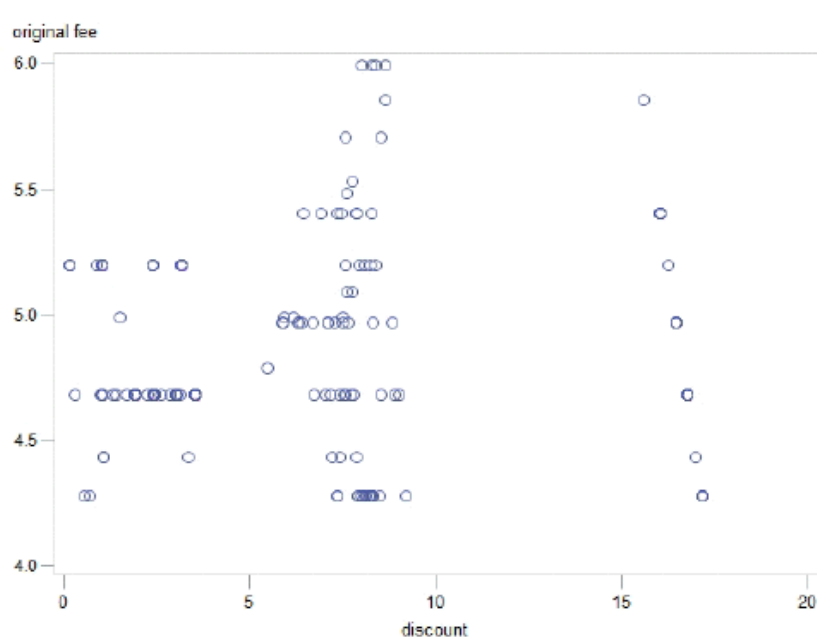
### ❖ Policy-violating bank representatives and negative discounts



*Distribution of negative discounts*

## Results Analysis (3/8)

### ❖ Policy-violating bank representatives and negative discounts



*Relationships between negative discounts and original and actual fees*

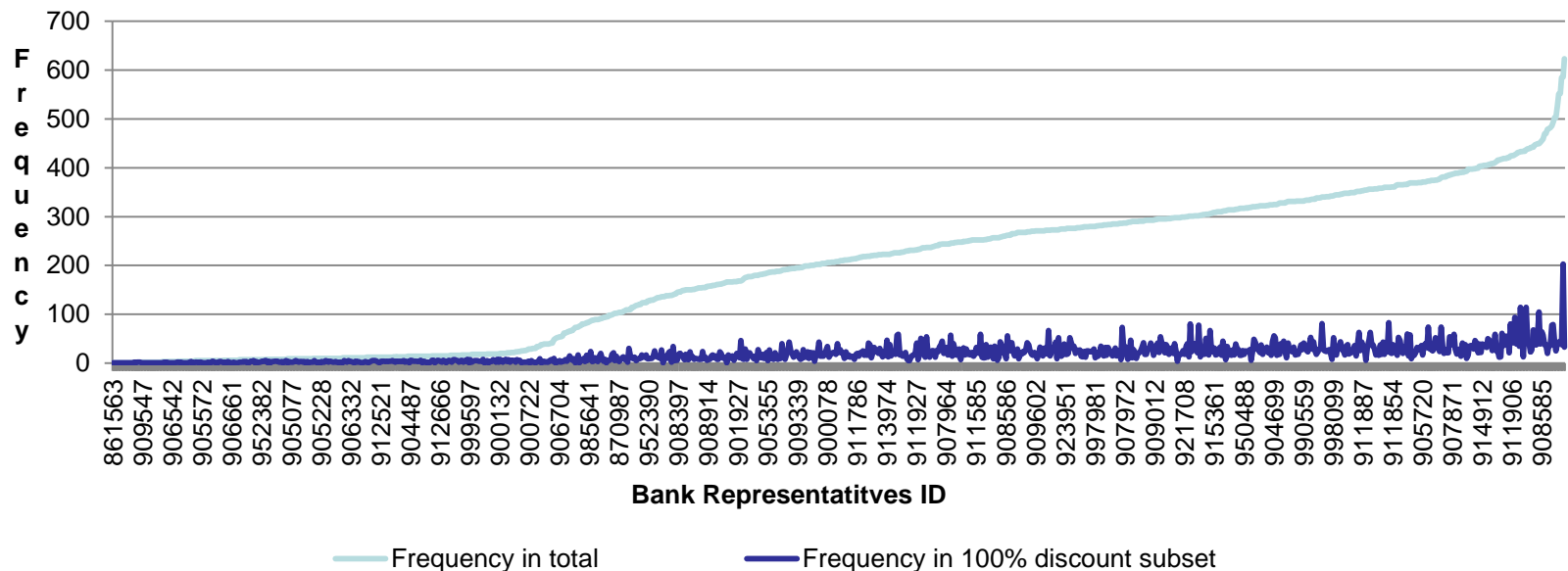
### ➤ New Audit Objective:

- ❖ Actual fees are recorded correctly.
- ❖ Original fees reflect the number of cards in an account.

# Results Analysis (4/8)

## ❖ Effortless bank representatives and inactive representatives

- Bank representatives who always offer 100% discounts should be considered not putting enough effort to negotiate with the clients for a lower discount.



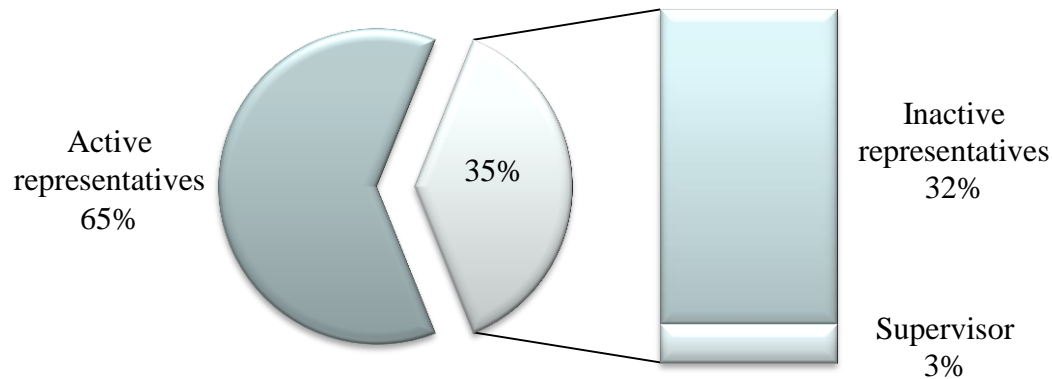
*Distribution of bank representatives offered 100% discounts in the whole retention data and the 100% discount subset*

## Results Analysis (5/8)

### ❖ Effortless bank representatives and inactive representatives

Mean	Standard Deviation	Minimum Value	Maximum Value	Count
170	148	1	623	1151

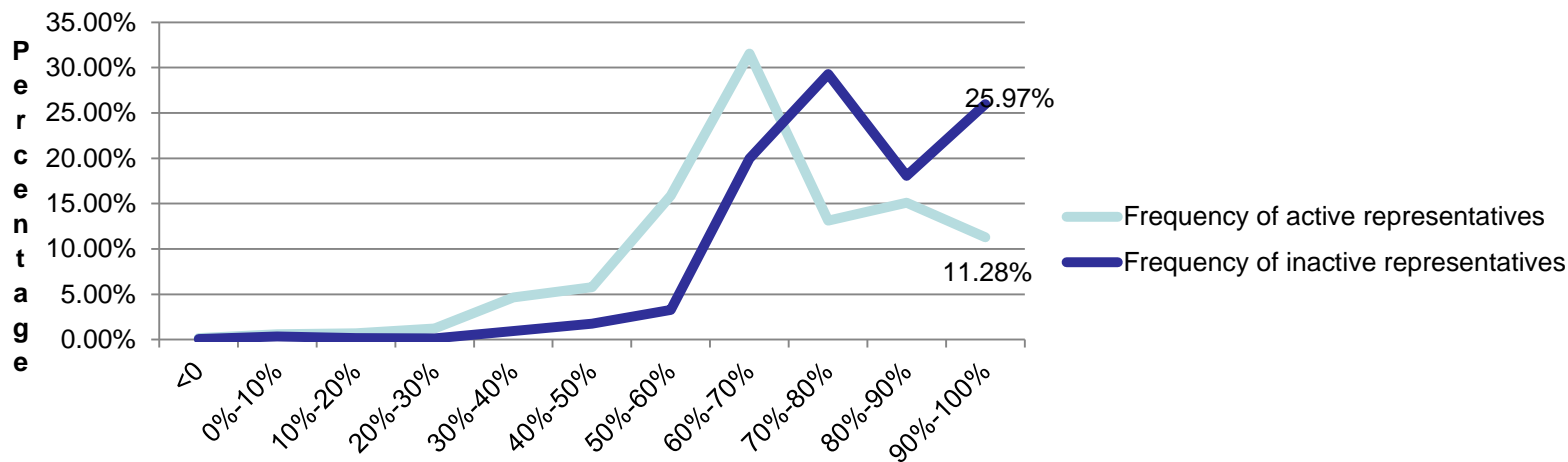
*Descriptive statistics of frequency distribution of bank representatives*



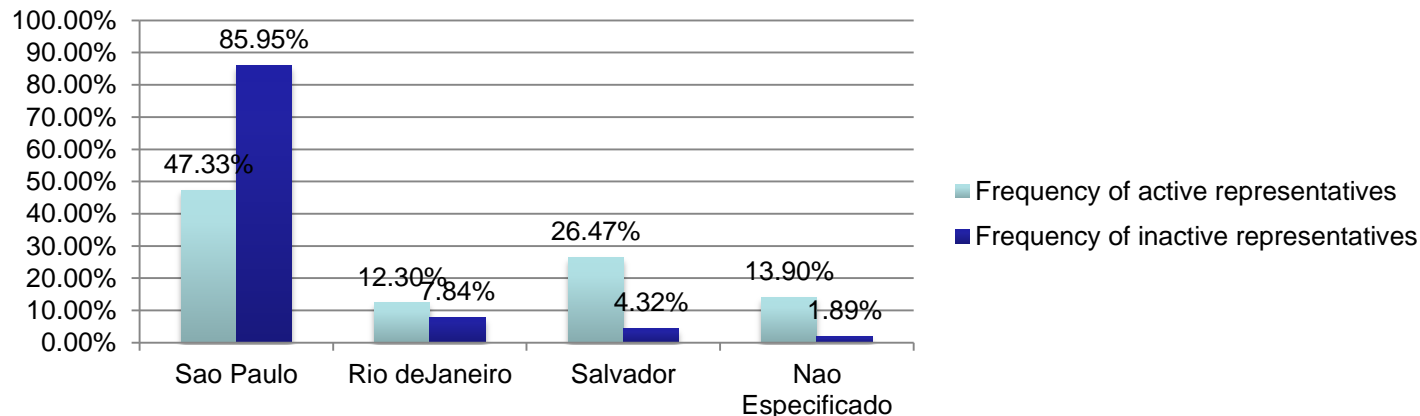
*Distribution of bank representatives*

# Results Analysis (6/8)

## ❖ Effortless bank representatives and inactive representatives



*Comparison of active and inactive representatives on frequency distribution of discounts*



*Distribution of bank representatives*



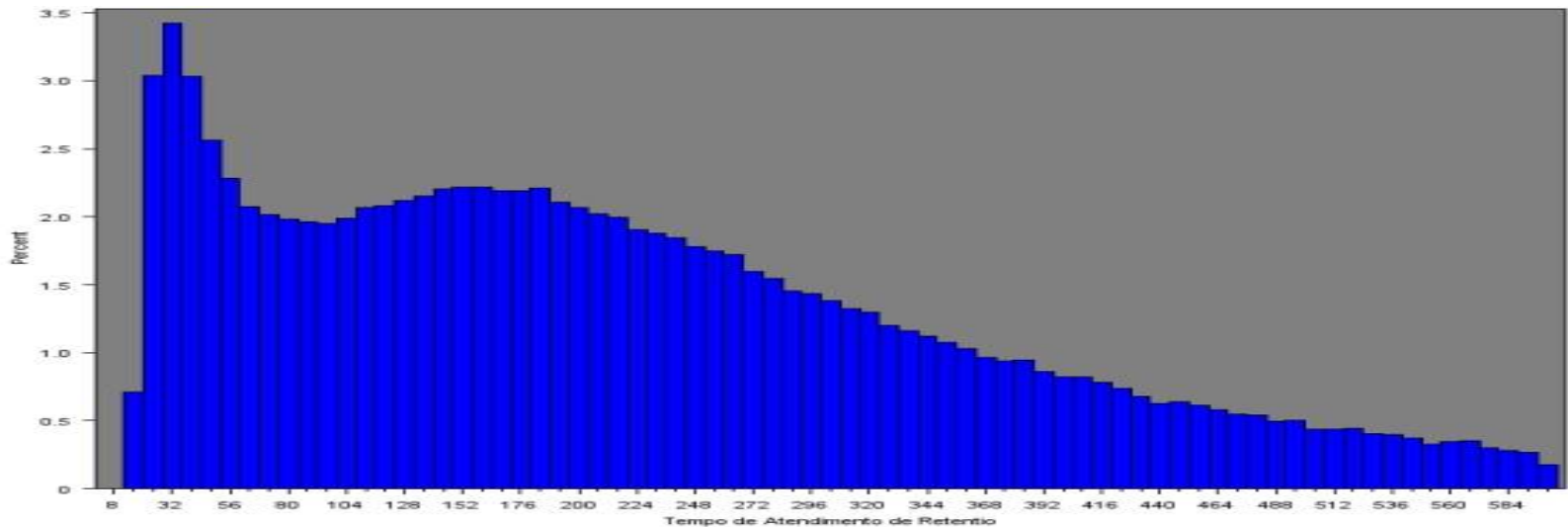
## Results Analysis (7/8)

### ❖ **Non-negotiation bank representatives and short calls**

- Bank representatives who offer a discount without negotiation usually related to short call duration

Minimum	Maximum	Mean	Median	90 <sup>th</sup> Percent
10	6561	255	206	514

*Descriptive statistics of call duration*

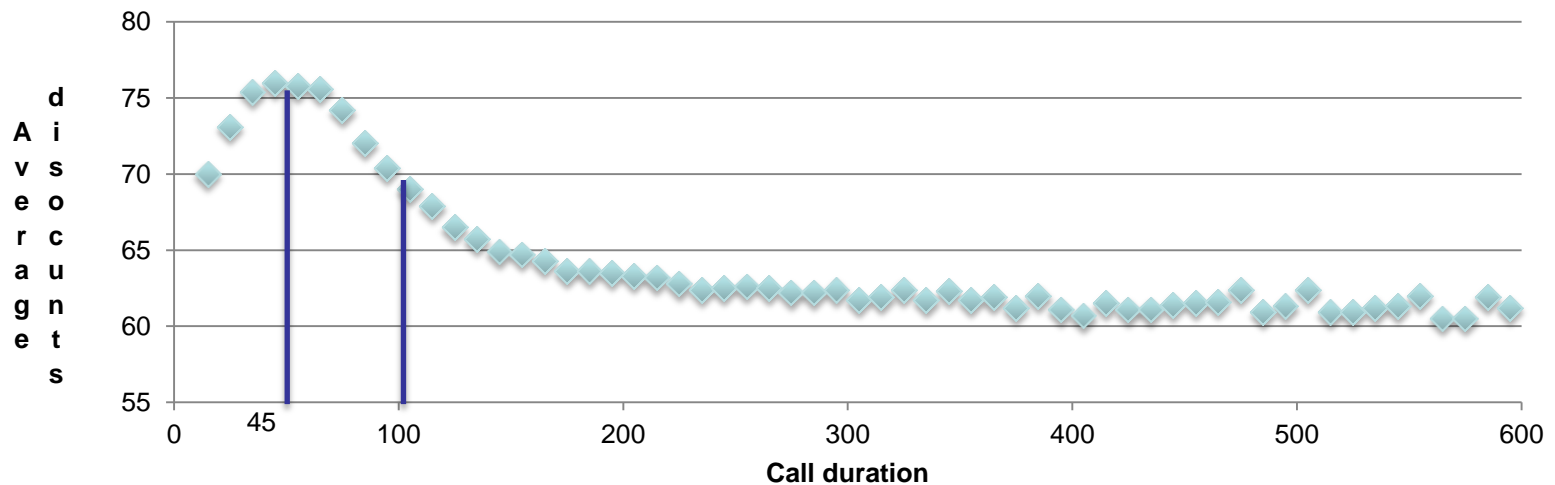


*Frequency distribution of call duration less than 600 seconds*

## Results Analysis (8/8)

### ❖ Non-negotiation bank representatives and short calls

- One possible explanation for these unreasonable short calls is that these calls are forced to terminate due to bad network connection.



*Relationship between call duration and discounts*

- New Audit Objective:

- ❖ All the discounts are given in calls long enough to offer discounts.

## Future research directions

- ❖ Demonstrate the application of EDA in the audit of financial statement related business cycle.
- ❖ Demonstrate the application of EDA in other types of auditing.
- ❖ Extend current framework to continuous auditing environment.
- ❖ Explore the application of other EDA technologies in auditing.
- ❖ Explore the most suitable EDA techniques for each audit procedure.

*Thank You!*