The Application of Exploratory Data Analysis (EDA) in Auditing

Qi Liu
Ph.D. Candidate (A.B.D.)
Dept. of Accounting & Information Systems
Rutgers University

28th 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).

<table>
<thead>
<tr>
<th>Reasoning Type</th>
<th>Exploratory Data Analysis (EDA)</th>
<th>Confirmatory Data Analysis (CDA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal</td>
<td>Pattern Recognition and Hypothesis generation</td>
<td>Estimation, Modeling, Hypothesis testing</td>
</tr>
<tr>
<td>Applied Data</td>
<td>Observation Data (data collected without well-defined hypothesis)</td>
<td>Experimental data (data collected through formally designed experiments)</td>
</tr>
<tr>
<td>Techniques</td>
<td>Descriptive Statistics, Data Visualization, Clustering Analysis, Process Mining…</td>
<td>Traditional statistical techniques of inference, significance, and confidence</td>
</tr>
<tr>
<td>Advantages</td>
<td>• No assumptions required</td>
<td>• Precise</td>
</tr>
<tr>
<td></td>
<td>• Promotes deeper understanding of the data</td>
<td>• Well-established theory and methods</td>
</tr>
<tr>
<td>Disadvantages</td>
<td>• No conclusive answers</td>
<td>• Required unrealistic assumptions</td>
</tr>
<tr>
<td></td>
<td>• Difficult to avoid bias produced by overfitting</td>
<td>• Difficult to notice unexpected results</td>
</tr>
</tbody>
</table>
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

Step 1
- Display the distribution of related fields

Step 2
- Identify salient features from the distribution

Step 3
- Perform CDA to test possible explanations

Step 4
- Identify suspicious cases
- Explore the causes of abnormal cases

Step 5
- Perform CDA to confirm the relationship

Step 6

Step 7
- Report the risks and recommend improvement suggestions
- Add a new audit objective
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

<table>
<thead>
<tr>
<th>Data (Retention Dataset)</th>
<th>Selected Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>➢ Each record represents a customer call</td>
<td>➢ Original fee (VLR_ANUIDADE_G)</td>
</tr>
<tr>
<td>➢ 195,694 records</td>
<td>➢ Actual fee (_Valor da Anuidade de Saída)</td>
</tr>
<tr>
<td>➢ 162 fields</td>
<td>➢ Agent identification (Funcional do Agente)</td>
</tr>
<tr>
<td>➢ Time frame: January, 2012</td>
<td>➢ Supervisor identification (Funcional do Supervisor)</td>
</tr>
<tr>
<td></td>
<td>➢ Location of the customer service center (Polo de Atendimento)</td>
</tr>
<tr>
<td></td>
<td>➢ Call duration (Tempo de Atendimento de Retenti)</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discount</td>
<td>-2.326.04</td>
<td>60</td>
<td>-27,944,522.22</td>
<td>100.00</td>
<td>219933.88</td>
</tr>
</tbody>
</table>

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

New Audit Objective:

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

Relationships between negative discounts and original and actual fees
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

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum Value</th>
<th>Maximum Value</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active representatives</td>
<td>170</td>
<td>148</td>
<td>1</td>
<td>623</td>
<td>1151</td>
</tr>
</tbody>
</table>

Descriptive statistics of frequency distribution of bank representatives

Distribution of bank representatives

Active representatives 65%

Inactive representatives 32%

Supervisor 3%
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

<table>
<thead>
<tr>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Median</th>
<th>90th Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>6561</td>
<td>255</td>
<td>206</td>
<td>514</td>
</tr>
</tbody>
</table>

Descriptive statistics of call duration

Frequency distribution of call duration less than 600 seconds
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.

Results Analysis (8/8)

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!