GHSUG User Group

The Means Procedure in SAS Enterprise Guide (EG)… “Summary Statistics”

Sueheir Saddik

April 15th, 2010
Outline

- The Means Procedure – Overview and Usage
- The Means Procedure – Proc Means Syntax
- SAS Enterprise Guide – Summary Statistics – Screen Shot and options
- Data Used and Information / Analysis Needed
- SAS Enterprise Guide – Summary Statistics
  - Data & Drop Down
  - Interface, Variable Selection, Statistics, Plots, Results, Titles, Code Viewing ... Etc
  - Past Due Amount Summary
  - Data / Code Checking
  - Make it Pretty & Add Distributions
  - Profiling (Key Variables), Output, Summary
- Questions
The Means Procedure – Overview and Usage

- The Means procedure is one of the most powerful and useful procedures in SAS

- The Means procedure computes descriptive statistics for variables across the observations and within groups of observation with an option to store the statistics in a SAS dataset
  - Some of the computed statistics are:
    - Arithmetic Mean: \( \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \)
    - Variance: \( Var(X) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} \)
    - Standard deviation: \( SD(X) = \sqrt{Var(X)} \)

- The Means Procedure can be used in; analytics, business intelligence, reporting, profiling, standard data checking for coded values created using exiting ones, checking for outliers / extreme values, missing values, data penetration, differences between population mean and certain group of observations (t-test) ... etc
The Means Procedure – Proc Means Syntax

- **Simplified ...**
  
  ```
  proc means data = XXXX mean min max n nmiss;
  var variable(s) name;
  run;
  ```

- **Complicated .... SAS web site**
  ```
  PROC MEANS <option(s)> <statistic-keyword(s)>;
  BY <DESCENDING> variable-1 <... <DESCENDING> variable-n><NOTSORTED>;
  CLASS variable(s) </ option(s)>;
  FREQ variable;
  ID variable(s);
  OUTPUT <OUT=SAS-data-set> <output-statistic-specification(s)> <id-group-specification(s)> <maximum-id-specification(s)> <minimum-id-specification(s)> </ option(s)> ;
  TYPES request(s);
  VAR variable(s) < / WEIGHT=weight-variable>;
  WAYS list;
  WEIGHT variable;
  ...
  RUN;
  ```

- **SAS EG is user friendly, with drag and drop interface. There is no need to remember or memorize the above code ... 😊**
  - More efficient
  - Less time consuming
  - Keeping it simple
SAS Enterprise Guide – Summary Statistics – Screen Shot
SAS Enterprise Guide – Summary Statistics

## Task Roles

- **Analysis variables:**
  - numeric variables on which you want the statistics generated

- **Classification variables:**
  - no data sorting required

- **Frequency count:**

- **Relative weight:**

- **Copy variables**

- **Group analysis by:**
  - requires that the data set must be sorted ... The variables that you assign to this role are used to compute separate statistics for each distinct value or combination of values of the Group analysis by variables. The data is automatically sorted by the variables in this role before the statistics are computed.

## Statistics

- **Basic**

- **Percentiles**

- **Additional**

## Plots

- **Histogram**

- **Box and Wisker**

## Results

- **Save statistics to data set**

- **Show statistics**

## Titles

- **Default**

- **specify**
Statistics

Basic statistics available:

- **Mean**: is the arithmetic average, calculated by adding the values of a sample variable and dividing this sum by the number of observations.
- **Std Dev = Standard deviation**: is a statistical measure of the variability of a group of data values. This measure, which is the most widely used measure of the dispersion of a frequency distribution, is equal to the positive square root of the variance.
- **Standard error**: is the standard deviation of the sample mean. The standard error is defined as the ratio of the sample standard deviation to the square root of the sample size.
- **Variance**: is a statistical measure of dispersion of data values. This measure is an average of the total squared dispersion between each observation and the sample mean.
- **Minimum**: is the lowest value for an observation.
- **Maximum**: is the largest value for an observation.
- **Range**: is the difference between the largest and the smallest values in the data.
- **Sum**: is the sum of all observations.
- **N = Number of observations**: is the total number of observations that do not have a missing value.
- **N Miss = Number of missing observations**: is the number of observations for which no value is entered.

- You can specify the maximum number of decimal places here
  - By default, a statistic is displayed by using the best fit, which is usually 7 decimal places.
- Missing values are not included in the calculations produced by the means procedure.
Statistics

- **Percentiles** … By default, no percentiles are selected.
  - 1st Pctl = 1st percentile
  - 5th Pctl = 5th percentile
  - 10th Pctl = 10th percentile
  - Lower Quartile = 1st quartile = 25\textsuperscript{th} percentile
  - Median = 50\textsuperscript{th} percentile
  - Upper Quartile = 3\textsuperscript{rd} quartile = 75\textsuperscript{th} percentile
  - 90th Pctl = 90th percentile
  - 95th Pctl = 95th percentile
  - 99th Pctl = 99th percentile

- **Additional**
  - Confidence limits of the mean
  - T – statistics and prob > |t| |
  - Coefficient of variation
  - Corrected sum of squares
  - Uncorrected sum of squares
Data Used and Information / Analysis Needed

- Bell Canada data was used in the example presented in this presentation
  - Collection data with 604,009 distinct observations
  - 74 variables ... Variables used
    - Past Due Amount: amount owing by customers
    - Bad_Ind_New: an indicator with these values
      - 0 = good
      - 1 = bad
    - Aging buckets
    - Time spent in collection in days
    - Tenure in months

- Information needed - questions to answer ....
  - Data quality, missing values, penetration, extreme values, transformations created ... ?
  - What is the overall average past due amount?
  - How does the past due amount differ with respect to the bad / good groups?
  - What is the profile of bad customers?
Data
Variables = columns
Observations = rows

Describe –
Summary Statistics

A value that exceeds 99% of the sample data values and is exceeded by 1% of the sample data values.
No plots will be shown today ... My advice, avoid the plots option unless you have super CPU ...
The process consumes lots and lots of CPU and brings the system down to it’s knees ...

Email received ... Oops ... subject “Fs: SAS Critical Alert - Server Conso SAS VGRCM1 CPU Util ...”
SAS Enterprise Guide – Summary Statistics – Results
SAS Enterprise Guide – Summary Statistics – Running & Done
### SAS output

**Analysis Variable:** PAST_DUE_AMT PAST_DUE_AMT

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Std Error</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Sum</th>
<th>N</th>
<th>Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>10th Pctl</th>
<th>Lower Quartile</th>
<th>Median</th>
<th>Upper Quartile</th>
<th>90th Pctl</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>45.3095348</td>
<td>105.5042062</td>
<td>0.1357526</td>
<td>-24929.890</td>
<td>4113.66</td>
<td>28517.59</td>
<td>57416163.59</td>
<td>604009</td>
<td>0</td>
<td>-93.67</td>
<td>-0.58</td>
<td>0</td>
<td>60.49</td>
<td>137.59</td>
<td>198.59</td>
<td>355.92</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Generated by the SAS System (SASMain, NET_ASIV) on 13APR2011 at 12:47 PM

### SAS output in a data set

<table>
<thead>
<tr>
<th><em>TYPE</em></th>
<th><em>FREQ</em></th>
<th>PAST_DUE_AMT Mean</th>
<th>PAST_DUE_AMT StdDev</th>
<th>PAST_DUE_AMT SdEn</th>
<th>PAST_DUE_AMT Min</th>
<th>PAST_DUE_AMT Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>UG</td>
<td>604009</td>
<td>45.3095348</td>
<td>105.5042062</td>
<td>0.1357526</td>
<td>-24929.890</td>
<td>4113.66</td>
</tr>
</tbody>
</table>

Bell Canada Restricted

![SAS Enterprise Guide interface](image)

**Summary Statistics for X_SS_11_SEGMENTATION**

- **Task Roles**
  - **Variables to assign:**
    - Name:
      - BAN
      - EXTRACT_DATE
      - BAN_STATUS
      - AGE_BUCKET_1_30_CURR
      - AGE_BUCKET_1_30_IND
      - AGE_BUCKET_31_60_CURR
      - AGE_BUCKET_31_60_IND
      - AGE_BUCKET_61_90_CURR
      - AGE_BUCKET_61_90_IND
      - AGE_BUCKET_91_120_CURR
      - AGE_BUCKET_91_120_IND
      - AGE_BUCKET_120_P_CURR
      - AGE_BUCKET_120_P_IND
      - AP_BALANCE
      - AP_BALANCE_Q
      - AREA
      - AUTO_GEN_PYM_TYPE
      - BILL_CYCLE
      - BILL_MEDIA
      - CHARGE_AMT
      - DISPUTE_AMT
      - LUMIFAS_ID (add to code)
  - **Task roles:**
    - Analyze variables
      - PAST_DUE_AMT
    - Classification variables
      - Frequency count (Limit: 1)
      - Relative weight (Limit: 1)
    - Copy variables
    - Group analysis by
      - Bad_Indicator_New

**Sort by variables:**

![Sort by variables](image)

**Variable to add option:**

![Variable to add option](image)
**Summary Statistics**

**Analysis Variable:** PAST_DUE_AMT PAST_DUE_AMT

<table>
<thead>
<tr>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Sum</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>10th Pctl</th>
<th>90th Pctl</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>33.4407213</td>
<td>10.020000756</td>
<td>-24583.93</td>
<td>3467.54</td>
<td>203770.0699</td>
<td>616561</td>
<td>0</td>
<td>-56.08</td>
<td>-0.73</td>
<td>0</td>
<td>0</td>
<td>64.16</td>
<td>122.17</td>
<td>174.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Sum</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>10th Pctl</th>
<th>90th Pctl</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>80.59665</td>
<td>118.0048121</td>
<td>-2231.95</td>
<td>4133.96</td>
<td>7065.91</td>
<td>7011016.49</td>
<td>87368</td>
<td>0</td>
<td>-20.79</td>
<td>0</td>
<td>0</td>
<td>60.27</td>
<td>114.38</td>
<td>228.18</td>
</tr>
</tbody>
</table>

Generated by the SAS System (SASMain, NET_ATsrv) on 14MAR2011 at 12:49 PM.
There are no missing past due amount values

604,009 observation
  - 516,651 good (86%)
  - 87,358 Bad (14%)

Bad rate = (87,358 / 604,009) = 14%

50% of all customers have no past due amount

50% of bad customers have a past due amount of $50 or less

Few extreme values /outliers (Min & Max ) such as (-$24,384, -$2,922, ..., $3,468, $4,134)

There average past due amount for good is significantly lower than that of bad
### Summary Statistics Interface

<table>
<thead>
<tr>
<th>Task roles:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis variables</td>
<td>![PAST_DUE_AMT G]</td>
</tr>
<tr>
<td>Classification variables</td>
<td>![PAST_DUE_AMT G]</td>
</tr>
<tr>
<td>Frequency count (Limit 1)</td>
<td>![PAST_DUE_AMT G]</td>
</tr>
<tr>
<td>Relative weight (Limit 1)</td>
<td>![PAST_DUE_AMT G]</td>
</tr>
<tr>
<td>Copy variables</td>
<td>![PAST_DUE_AMT G]</td>
</tr>
<tr>
<td>Group analysis by</td>
<td>![PAST_DUE_AMT G]</td>
</tr>
</tbody>
</table>

#### SAS Code created

```sas
if PAST_DUE_AMT < 0 then PAST_DUE_AMT_G = '-ive
else if PAST_DUE_AMT = 0 then PAST_DUE_AMT_G = '$0
else if 0 < PAST_DUE_AMT <= 75 then PAST_DUE_AMT_G = '1-$0.01 to $75
else if 75 < PAST_DUE_AMT <= 150 then PAST_DUE_AMT_G = '2-$75.01 to $150
else if 150 < PAST_DUE_AMT <= 250 then PAST_DUE_AMT_G = '3-$150.01 to $250
else if 250 < PAST_DUE_AMT <= 500 then PAST_DUE_AMT_G = '4-$250.01 to $500
else if PAST_DUE_AMT > 500 then PAST_DUE_AMT_G = '5-$500.01 +
run;
```

### SAS output

#### PAST_DUE_AMT_G=$0

<table>
<thead>
<tr>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>Median</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>276403</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

#### PAST_DUE_AMT_G=-ive

<table>
<thead>
<tr>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>Median</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>-32.2243993</td>
<td>-24383.93</td>
<td>-0.01</td>
<td>36913</td>
<td>0</td>
<td>-343.11</td>
<td>-100.3</td>
<td>-3.86</td>
<td>-0.09</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

#### PAST_DUE_AMT_G=1-$0.01 to $75

<table>
<thead>
<tr>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>Median</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>46.6463012</td>
<td>0.01</td>
<td>75</td>
<td>160903</td>
<td>0</td>
<td>0.1</td>
<td>2.75</td>
<td>50.94</td>
<td>72.62</td>
<td>74.56</td>
</tr>
</tbody>
</table>

#### PAST_DUE_AMT_G=2-$75.01 to $150

<table>
<thead>
<tr>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>Median</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>106.2728741</td>
<td>75.01</td>
<td>150</td>
<td>77476</td>
<td>0</td>
<td>75.49</td>
<td>77.38</td>
<td>103.14</td>
<td>143.95</td>
<td>148.91</td>
</tr>
</tbody>
</table>

#### PAST_DUE_AMT_G=3-$150.01 to $250

<table>
<thead>
<tr>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>Median</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>190.4260796</td>
<td>150.01</td>
<td>250</td>
<td>35405</td>
<td>0</td>
<td>150.63</td>
<td>153.11</td>
<td>186.4</td>
<td>240.87</td>
<td>248.08</td>
</tr>
</tbody>
</table>

#### PAST_DUE_AMT_G=4-$250.01 to $500

<table>
<thead>
<tr>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>Median</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>326.9079031</td>
<td>250.01</td>
<td>500</td>
<td>15094</td>
<td>0</td>
<td>250.72</td>
<td>254.11</td>
<td>308.815</td>
<td>459.35</td>
<td>491.93</td>
</tr>
</tbody>
</table>

#### PAST_DUE_AMT_G=5-$500.01 +

<table>
<thead>
<tr>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>Median</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>656.8003251</td>
<td>500.05</td>
<td>4133.66</td>
<td>1815</td>
<td>0</td>
<td>502.24</td>
<td>507.38</td>
<td>591.94</td>
<td>995.9</td>
<td>1435.91</td>
</tr>
</tbody>
</table>
### Data Checking Output - Make it Pretty & Add Distributions

**Analysis Variable:** PAST_DUE_AMT

<table>
<thead>
<tr>
<th>PAST_DUE_AMT_G</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>Median</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
<th>Distribution</th>
<th>Cumulative Distribution</th>
<th>Inverse Cumulative Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>276,403</td>
<td>0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>46%</td>
<td>46%</td>
<td>100%</td>
</tr>
<tr>
<td>-ive</td>
<td>-$32</td>
<td>-$24,384</td>
<td>-$0</td>
<td>36,913</td>
<td>0</td>
<td>-$343</td>
<td>-$100</td>
<td>-$4</td>
<td>-$0</td>
<td>-$0</td>
<td>6%</td>
<td>52%</td>
<td>54%</td>
</tr>
<tr>
<td>1-$0.01 to $75</td>
<td>$47</td>
<td>$0</td>
<td>$75</td>
<td>160,903</td>
<td>0</td>
<td>$0</td>
<td>$3</td>
<td>$51</td>
<td>$73</td>
<td>$75</td>
<td>27%</td>
<td>79%</td>
<td>48%</td>
</tr>
<tr>
<td>2-$75.01 to $150</td>
<td>$106</td>
<td>$75</td>
<td>$150</td>
<td>77,476</td>
<td>0</td>
<td>$75</td>
<td>$77</td>
<td>$103</td>
<td>$144</td>
<td>$149</td>
<td>13%</td>
<td>91%</td>
<td>21%</td>
</tr>
<tr>
<td>3-$150.01 to $250</td>
<td>$190</td>
<td>$150</td>
<td>$250</td>
<td>35,405</td>
<td>0</td>
<td>$151</td>
<td>$153</td>
<td>$186</td>
<td>$241</td>
<td>$248</td>
<td>6%</td>
<td>97%</td>
<td>9%</td>
</tr>
<tr>
<td>4-$250.01 to $500</td>
<td>$327</td>
<td>$250</td>
<td>$500</td>
<td>15,094</td>
<td>0</td>
<td>$251</td>
<td>$254</td>
<td>$309</td>
<td>$459</td>
<td>$492</td>
<td>2%</td>
<td>100%</td>
<td>3%</td>
</tr>
<tr>
<td>5-$500.01 +</td>
<td>$657</td>
<td>$500</td>
<td>$4,134</td>
<td>1,815</td>
<td>0</td>
<td>$502</td>
<td>$507</td>
<td>$592</td>
<td>$996</td>
<td>$1,436</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>$45</td>
<td>-$24,384</td>
<td>$4,134</td>
<td>604,009</td>
<td>0</td>
<td>-$34</td>
<td>-$1</td>
<td>$0</td>
<td>$199</td>
<td>$359</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Data created is correct ...
- 46% of customers have no past due amount
- 6% of customer have credits
- 79% of all customers have at most $75 past due amount
- 21% of customers have at least $75 past due amount
SAS Enterprise Guide – Summary Statistics – Profiling (Key Variables)
### SAS Enterprise Guide – Summary Statistics – Profiling Output

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Sum</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>10th Pctl</th>
<th>90th Pctl</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAST_DUE_AMT</td>
<td>39.2</td>
<td>102.8</td>
<td>-240.4</td>
<td>9468</td>
<td>11675</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-36.0</td>
<td>-1.0</td>
<td>0.0</td>
<td>64.0</td>
<td>122.0</td>
<td>175.0</td>
</tr>
<tr>
<td>CHARGE_AMT</td>
<td>127</td>
<td>134.0</td>
<td>0</td>
<td>57886</td>
<td>57886</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>24.0</td>
<td>38.0</td>
<td>47.0</td>
<td>72.0</td>
<td>102.0</td>
<td>155.0</td>
</tr>
<tr>
<td>DISPUTE_AMT</td>
<td>0</td>
<td>1.0</td>
<td>-400.0</td>
<td>060.6</td>
<td>050.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>AGE_BUCKET_1_30_CURR</td>
<td>25.4</td>
<td>47.0</td>
<td>0</td>
<td>012.9</td>
<td>012.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>AGE_BUCKET_31_60_CURR</td>
<td>4.7</td>
<td>16.0</td>
<td>0</td>
<td>082.4</td>
<td>082.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>AGE_BUCKET_61_90_CURR</td>
<td>1.6</td>
<td>6.0</td>
<td>0</td>
<td>099.4</td>
<td>099.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>AGE_BUCKET_91_120_CURR</td>
<td>0</td>
<td>2.0</td>
<td>0</td>
<td>054.0</td>
<td>054.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>AGE_BUCKET_120_P_CURR</td>
<td>0</td>
<td>2.0</td>
<td>0</td>
<td>032.1</td>
<td>032.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>MONTHS_SINCE_OLEST_RA</td>
<td>132</td>
<td>119.0</td>
<td>0</td>
<td>039.0</td>
<td>039.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Sum</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>10th Pctl</th>
<th>90th Pctl</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAST_DUE_AMT</td>
<td>81.0</td>
<td>119.0</td>
<td>-2022.0</td>
<td>4124.0</td>
<td>6146.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-21.0</td>
<td>0.0</td>
<td>0.0</td>
<td>60.0</td>
<td>114.0</td>
<td>226.0</td>
</tr>
<tr>
<td>CHARGE_AMT</td>
<td>185.0</td>
<td>189.0</td>
<td>0</td>
<td>15140.0</td>
<td>15140</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26.0</td>
<td>41.0</td>
<td>51.0</td>
<td>78.0</td>
<td>141.0</td>
<td>238.0</td>
</tr>
<tr>
<td>DISPUTE_AMT</td>
<td>0</td>
<td>2.0</td>
<td>-387.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>AGE_BUCKET_1_30_CURR</td>
<td>44.0</td>
<td>69.0</td>
<td>0</td>
<td>082.0</td>
<td>082.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>AGE_BUCKET_31_60_CURR</td>
<td>13.0</td>
<td>37.0</td>
<td>0</td>
<td>1477.0</td>
<td>1477.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>AGE_BUCKET_61_90_CURR</td>
<td>2</td>
<td>14.0</td>
<td>0</td>
<td>022.0</td>
<td>022.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>AGE_BUCKET_91_120_CURR</td>
<td>0</td>
<td>5.0</td>
<td>0</td>
<td>049.0</td>
<td>049.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>AGE_BUCKET_120_P_CURR</td>
<td>0</td>
<td>5.0</td>
<td>0</td>
<td>089.0</td>
<td>089.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>MONTHS_SINCE_OLEST_RA</td>
<td>32.0</td>
<td>61.0</td>
<td>0</td>
<td>035.0</td>
<td>035.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>
### Bad customers profile

- Have significantly higher past due amount
- Have significantly higher charge amount
- Have significantly higher $$ in age bucket 1
- Tend to have low tenure
  - New customers!

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Sum</th>
<th>N</th>
<th>N Miss</th>
<th>1st Pctl</th>
<th>5th Pctl</th>
<th>10th Pctl</th>
<th>Lower Quartile</th>
<th>Median</th>
<th>Upper Quartile</th>
<th>90th Pctl</th>
<th>95th Pctl</th>
<th>99th Pctl</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAST_DUE_AMT</td>
<td>$39</td>
<td>102</td>
<td>-24,384</td>
<td>$3,468</td>
<td>27,851</td>
<td>20,377,088</td>
<td>516,651</td>
<td>0</td>
<td>-36</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>64</td>
<td>122</td>
<td>175</td>
<td>306</td>
</tr>
<tr>
<td>CHARGE_AMT</td>
<td>$127</td>
<td>134</td>
<td>0</td>
<td>$57,886</td>
<td>57,886</td>
<td>$65,719,403</td>
<td>516,521</td>
<td>130</td>
<td>$24</td>
<td>$38</td>
<td>$47</td>
<td>$72</td>
<td>$102</td>
<td>$155</td>
<td>$226</td>
<td>$286</td>
<td>$491</td>
</tr>
<tr>
<td>AGE_BUCKET_1</td>
<td>$25</td>
<td>47</td>
<td>0</td>
<td>$1,829</td>
<td>1,829</td>
<td>$12,805,045</td>
<td>516,521</td>
<td>130</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$5</td>
<td>$28</td>
<td>$83</td>
</tr>
<tr>
<td>AGE_BUCKET_2</td>
<td>$4</td>
<td>17</td>
<td>0</td>
<td>$2,088,773</td>
<td>824</td>
<td>$2,088,773</td>
<td>516,521</td>
<td>130</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$5</td>
<td>$28</td>
<td>$83</td>
</tr>
<tr>
<td>AGE_BUCKET_3</td>
<td>$1</td>
<td>6</td>
<td>0</td>
<td>$994</td>
<td>994</td>
<td>$330,437</td>
<td>516,521</td>
<td>130</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$21</td>
</tr>
<tr>
<td>AGE_BUCKET_4</td>
<td>$0</td>
<td>2</td>
<td>0</td>
<td>$67,627</td>
<td>454</td>
<td>$67,627</td>
<td>516,521</td>
<td>130</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$2</td>
</tr>
<tr>
<td>AGE_BUCKET_5</td>
<td>$0</td>
<td>2</td>
<td>0</td>
<td>$64,620</td>
<td>321</td>
<td>$64,620</td>
<td>516,521</td>
<td>130</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
</tr>
<tr>
<td>Tenure</td>
<td>132</td>
<td>116</td>
<td>0</td>
<td>359</td>
<td>359</td>
<td>68,067,605</td>
<td>516,651</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>33</td>
<td>99</td>
<td>214</td>
<td>351</td>
<td>354</td>
<td>356</td>
</tr>
<tr>
<td>PAST_DUE_AMT</td>
<td>$81</td>
<td>118</td>
<td>-2,922</td>
<td>$4,134</td>
<td>7,056</td>
<td>7,041,015</td>
<td>87,358</td>
<td>0</td>
<td>-21</td>
<td>0</td>
<td>0</td>
<td>50</td>
<td>114</td>
<td>226</td>
<td>308</td>
<td>515</td>
<td></td>
</tr>
<tr>
<td>CHARGE_AMT</td>
<td>$185</td>
<td>168</td>
<td>0</td>
<td>$15,140</td>
<td>15,140</td>
<td>$16,129,541</td>
<td>87,328</td>
<td>30</td>
<td>$26</td>
<td>$41</td>
<td>$51</td>
<td>$78</td>
<td>$141</td>
<td>$238</td>
<td>$364</td>
<td>$481</td>
<td>$761</td>
</tr>
<tr>
<td>AGE_BUCKET_1</td>
<td>$44</td>
<td>69</td>
<td>0</td>
<td>$3,826</td>
<td>3,826</td>
<td>$3,862,311</td>
<td>87,328</td>
<td>30</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$10</td>
<td>$131</td>
<td>$184</td>
</tr>
<tr>
<td>AGE_BUCKET_2</td>
<td>$13</td>
<td>37</td>
<td>0</td>
<td>$1,477</td>
<td>1,477</td>
<td>$1,091,917</td>
<td>87,328</td>
<td>30</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$185</td>
</tr>
<tr>
<td>AGE_BUCKET_3</td>
<td>$2</td>
<td>14</td>
<td>0</td>
<td>$622</td>
<td>622</td>
<td>$190,703</td>
<td>87,328</td>
<td>30</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$62</td>
</tr>
<tr>
<td>AGE_BUCKET_4</td>
<td>$0</td>
<td>5</td>
<td>0</td>
<td>$439</td>
<td>439</td>
<td>$30,289</td>
<td>87,328</td>
<td>30</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$8</td>
</tr>
<tr>
<td>AGE_BUCKET_5</td>
<td>$0</td>
<td>5</td>
<td>0</td>
<td>$688</td>
<td>688</td>
<td>$23,557</td>
<td>87,328</td>
<td>30</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$1</td>
</tr>
<tr>
<td>Tenure</td>
<td>32</td>
<td>61</td>
<td>0</td>
<td>356</td>
<td>356</td>
<td>2,814,915</td>
<td>87,358</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>31</td>
<td>94</td>
<td>155</td>
<td>351</td>
</tr>
<tr>
<td>PAST_DUE_AMT</td>
<td>$45</td>
<td>106</td>
<td>-24,384</td>
<td>$4,134</td>
<td>28,518</td>
<td>27,418,104</td>
<td>604,009</td>
<td>0</td>
<td>-34</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>68</td>
<td>138</td>
<td>199</td>
<td>359</td>
</tr>
<tr>
<td>CHARGE_AMT</td>
<td>$136</td>
<td>141</td>
<td>0</td>
<td>$57,886</td>
<td>57,886</td>
<td>$81,848,943</td>
<td>603,849</td>
<td>160</td>
<td>$24</td>
<td>$39</td>
<td>$48</td>
<td>$72</td>
<td>$105</td>
<td>$166</td>
<td>$247</td>
<td>$320</td>
<td>$563</td>
</tr>
<tr>
<td>AGE_BUCKET_1</td>
<td>$28</td>
<td>51</td>
<td>0</td>
<td>$3,826</td>
<td>3,826</td>
<td>$16,667,716</td>
<td>603,849</td>
<td>160</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$45</td>
<td>$82</td>
<td>$125</td>
</tr>
<tr>
<td>AGE_BUCKET_2</td>
<td>$5</td>
<td>22</td>
<td>0</td>
<td>$1,477</td>
<td>1,477</td>
<td>$3,180,691</td>
<td>603,849</td>
<td>160</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$11</td>
<td>$34</td>
<td>$106</td>
</tr>
<tr>
<td>AGE_BUCKET_3</td>
<td>$1</td>
<td>8</td>
<td>0</td>
<td>$994</td>
<td>994</td>
<td>$521,141</td>
<td>603,849</td>
<td>160</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$25</td>
</tr>
<tr>
<td>AGE_BUCKET_4</td>
<td>$0</td>
<td>3</td>
<td>0</td>
<td>$454</td>
<td>454</td>
<td>$97,916</td>
<td>603,849</td>
<td>160</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$3</td>
</tr>
<tr>
<td>AGE_BUCKET_5</td>
<td>$0</td>
<td>3</td>
<td>0</td>
<td>$688</td>
<td>688</td>
<td>$88,178</td>
<td>603,849</td>
<td>160</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
</tr>
<tr>
<td>Tenure</td>
<td>117</td>
<td>115</td>
<td>0</td>
<td>359</td>
<td>359</td>
<td>70,882,520</td>
<td>604,009</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>19</td>
<td>80</td>
<td>189</td>
<td>337</td>
<td>353</td>
<td>355</td>
</tr>
</tbody>
</table>
Questions ?