LENDING CLUB DATA ANALYSIS-ML

Applying machine learning models to the data and deploying best model in Azure



**Pranay Mankad**

**Priyesh Kaushik**

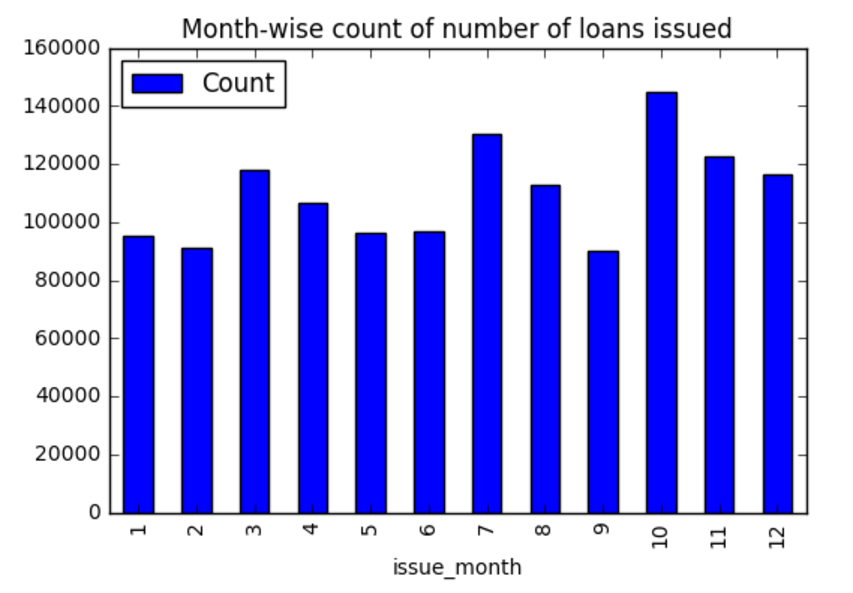
**Exploratory data analysis in Python**

**Analysis 1:** Find the month-wise count of number of loans issued

**Approach:**

Group the above dataframe for every month and count the number of loans issued in each month.

The result is saved in Analysis\_1.csv and a graph is drawn which represent this flow.



**Conclusion:** This graph shows us the month wise distribution on the number of loan issued.

**Analysis 2:** State-wise count of number of loans issued

**Approach:**

Group the dataframe for every state and count the number of loans issued in each state.

The result is saved in Analysis\_2.csv and a graph is drawn which represent this flow.

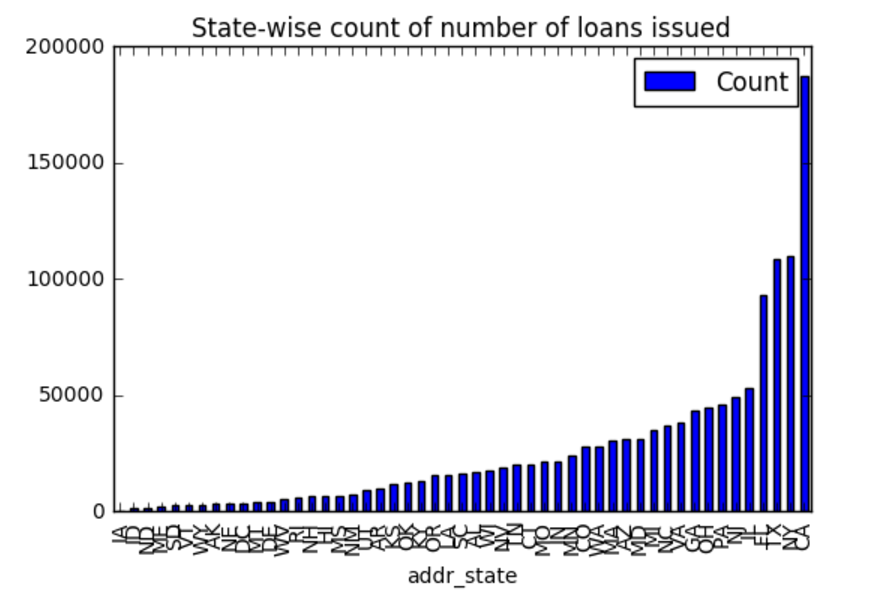
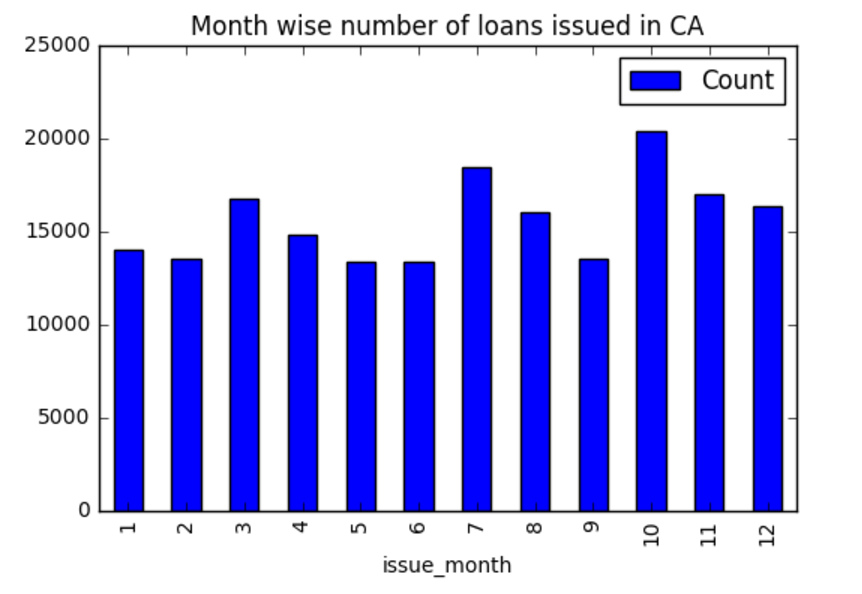
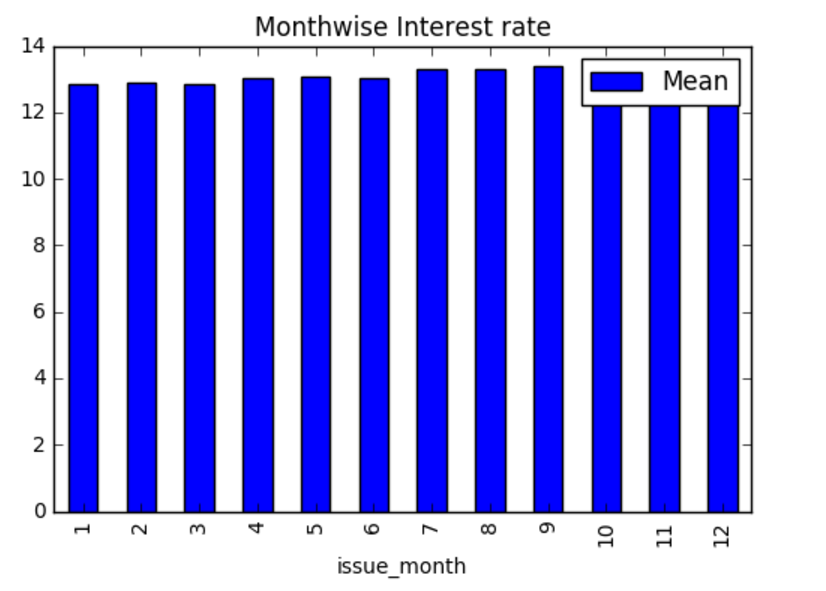
****

Figure 3: State wise loan issued distribution.

Here we see that CA has the highest count compared to other states and hence we dive into CA state issuers. On getting a month-wise graph, we observe that it is similar to the overall month-wise distribution.



Since there is no anomalies or sudden rise or fall in the graph we extract the interest rate in CA for each month.



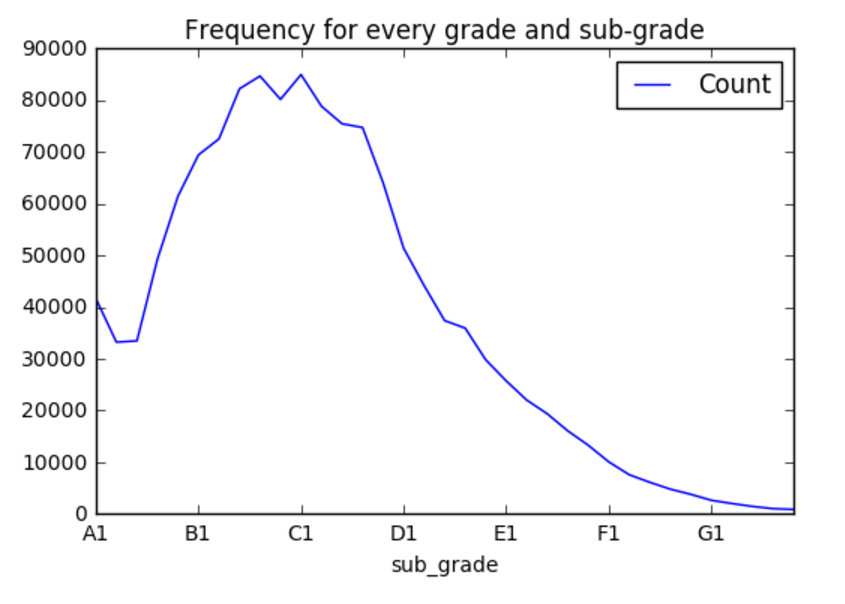
**Conclusion:** We observe that even though the interest rate is increasing over the months, there seems to be a rise in the number of people who are issued loans. This concludes that the Lending Club is targeting the right market in CA.

**Analysis 3:** Find the frequency for every grade and sub-grade

**Approach:**

Group the above dataframe for every grade and sub-grade and count the number of loans issued.

The result is saved in Analysis\_3.csv and a graph is drawn which represent this flow.

****

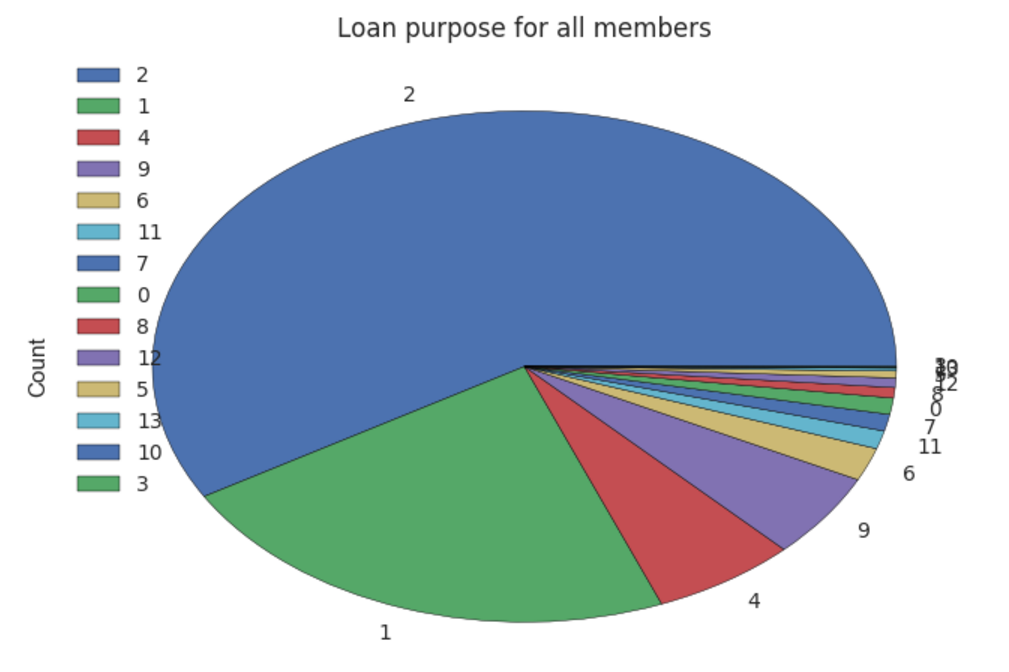
**Conclusion:** We observe that the maximum number of loan issuers lie in grades B3, B4, B5, C1 and C2.

**Analysis 4:** Find the loan purpose for all members

**Approach:**

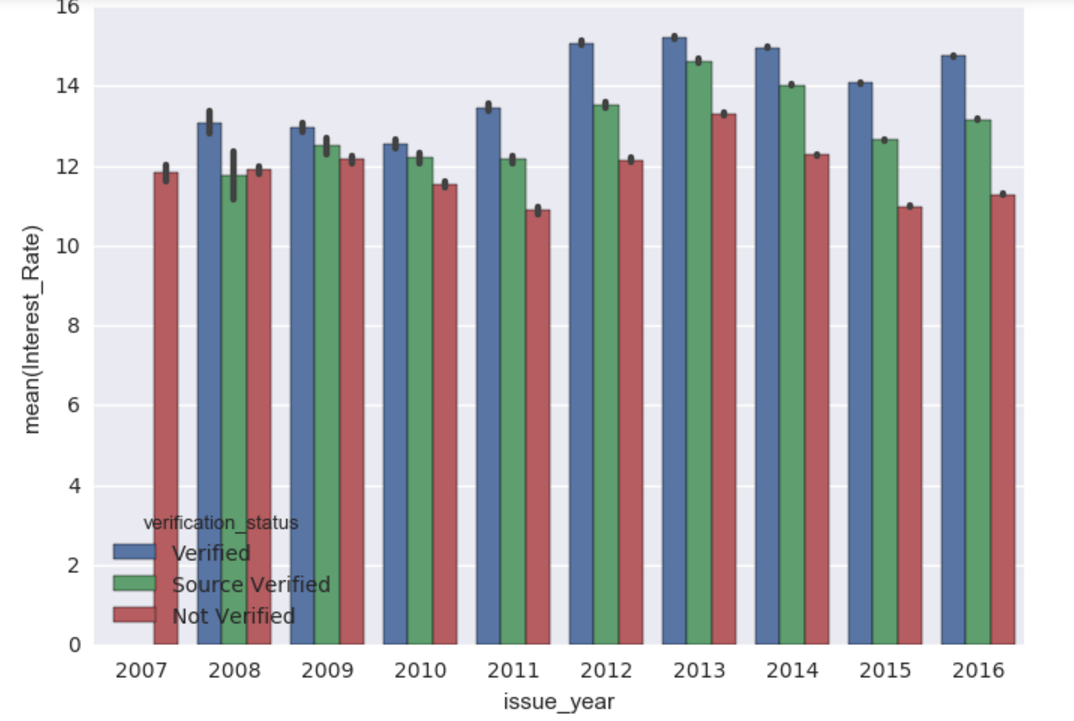
Group the above dataframe for every purpose and count the number of loans issued.

The result is saved in Analysis\_4.csv and a graph is drawn which represent this flow.



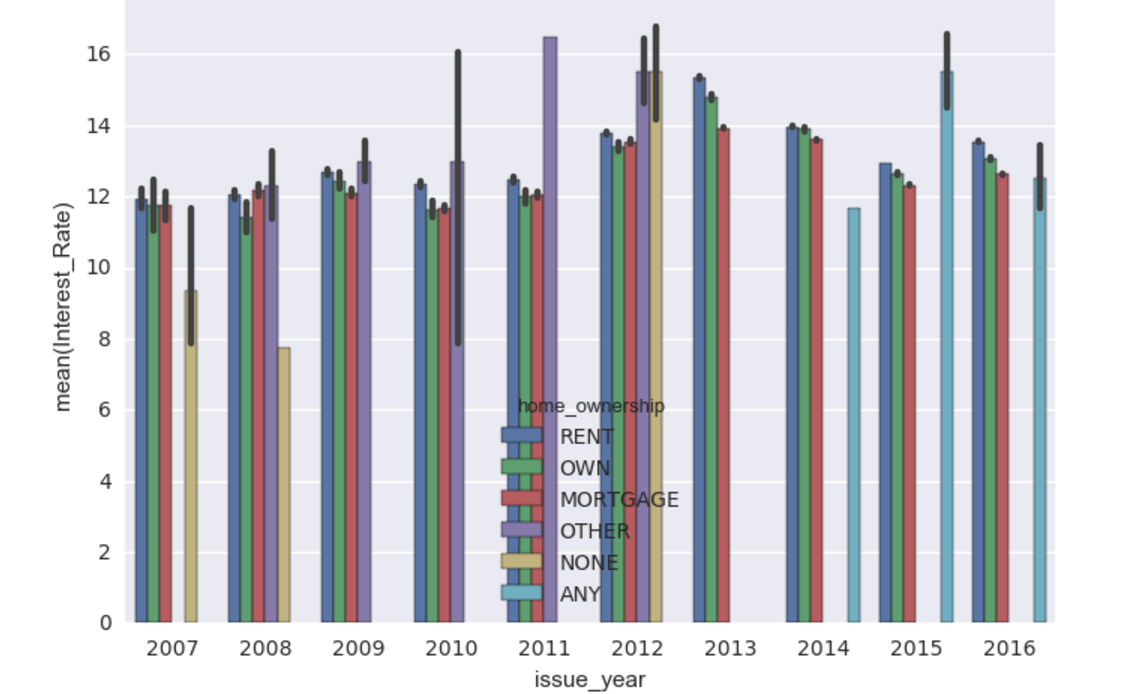
**Conclusion:** We observe that the most mentioned reason for issuing a loan is debt consolidation which is referred by the number 2.

**Analysis 5:** Find the mean Interest Rate for every Verification status over all the years.

****

**Conclusion:** We observe that the difference between various groups of Verification Status grows over the years from 2007 to 2016. Hence we can conclude that the verification team is working efficiently and processing more

**Analysis 6:** House ownership status for all members

****

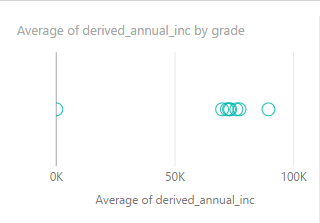
**Conclusion:** This graph represents the home ownership distribution for each year from 2007 to 2016. As the mean interest rate for people renting, owning and mortgaging a house is almost the same across all the years, we can conclude that this attribute does not make a huge difference to the credit score and interest rate.

## **Exploratory data analysis in Power BI**

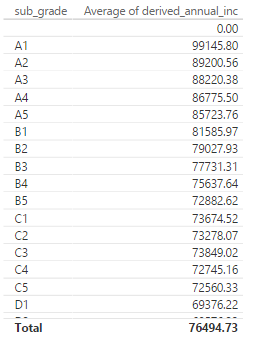
Once we have extracted the data, cleaned and pre-processed it, analyzed the data, completed the feature selection we come to the final task of making the data insights visually appealing using PowerBI.

Analysis is done for two reasons –

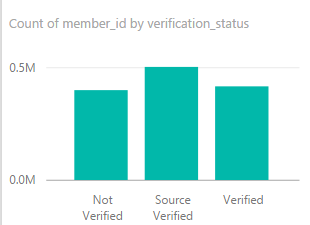
1. to get some favorable insights
2. for statistical purpose
3. Average annual income based on the Grade a person belongs to.
   * This gives the data of the people having an average salary based on the Grades they belong. By looking at the data, we get to know that only Grade A people are class apart and have way better salary than any other Grades. Any Grade other than A, has more or less a similar average salary.



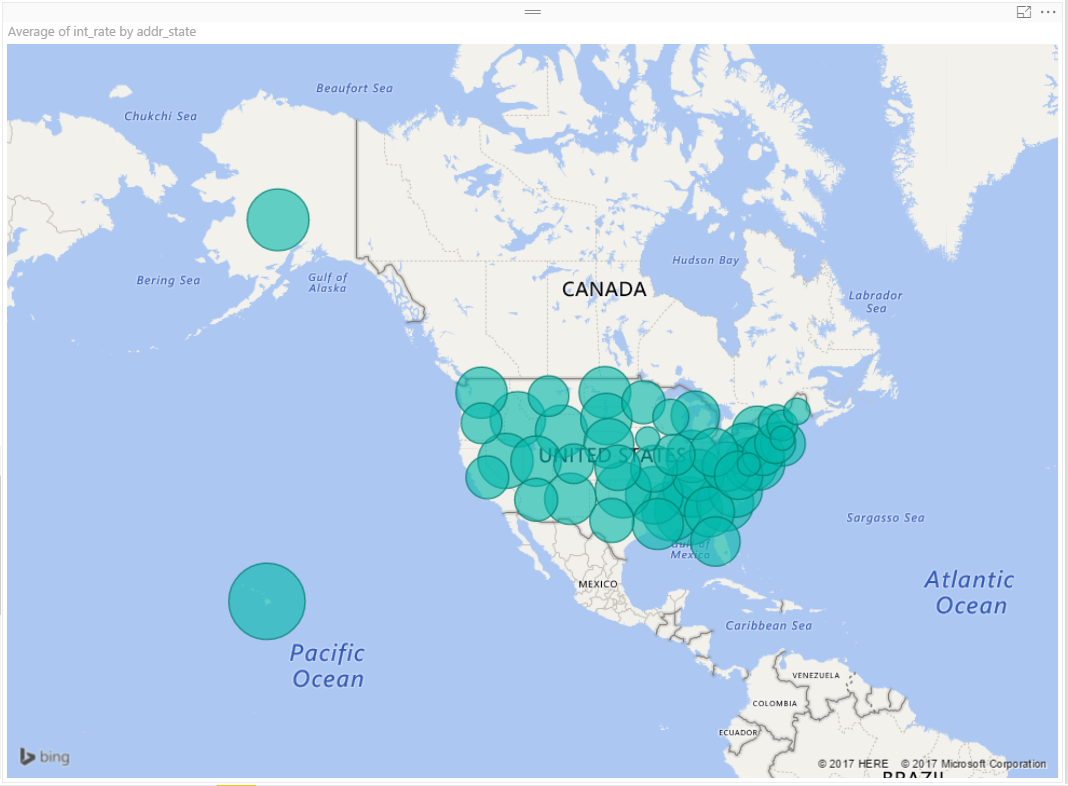
1. Getting even deeper and drilling down the hierarchy, the average salary based on the sub-grades of an individual.



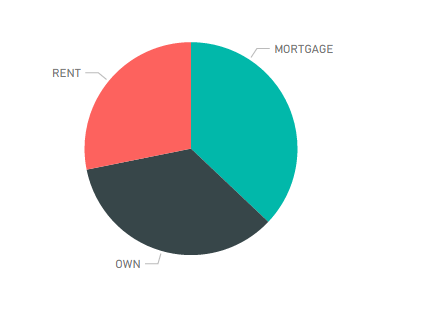
1. Statistical data on count of verified, not verified and Source verified individuals.



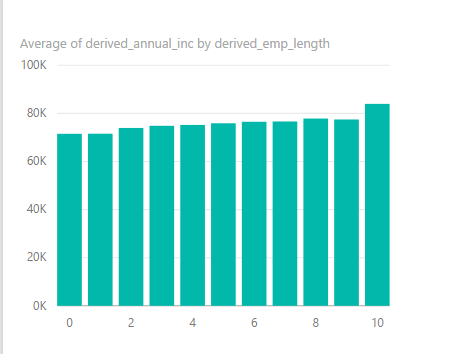
1. The average interest rate based on each state. On visualizing the data from the given link, we concluded that the closer the western part of the States has comparatively less interest rates.



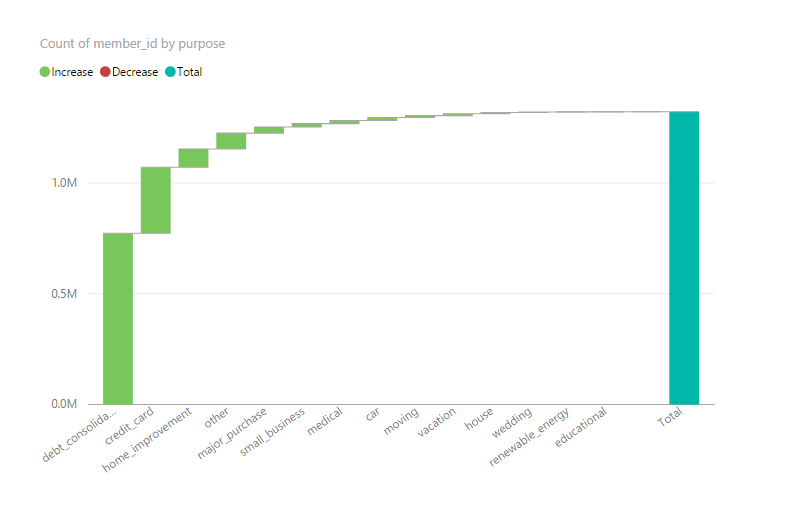
1. The average employee work length to acquire different types of home ownership.



1. Salary to number of years a person has worked. This shows that salary increases with amount of experience.



1. This analysis is probably the most interesting which helps us find the most popular reason for a person to land into Lender’s Club to borrow money.



PowerBI dashboards are available in the below link:

<https://app.powerbi.com/groups/me/reports/6a3b7b14-1a9c-44b2-8cf5-9a7db45bb3bf/ReportSection>