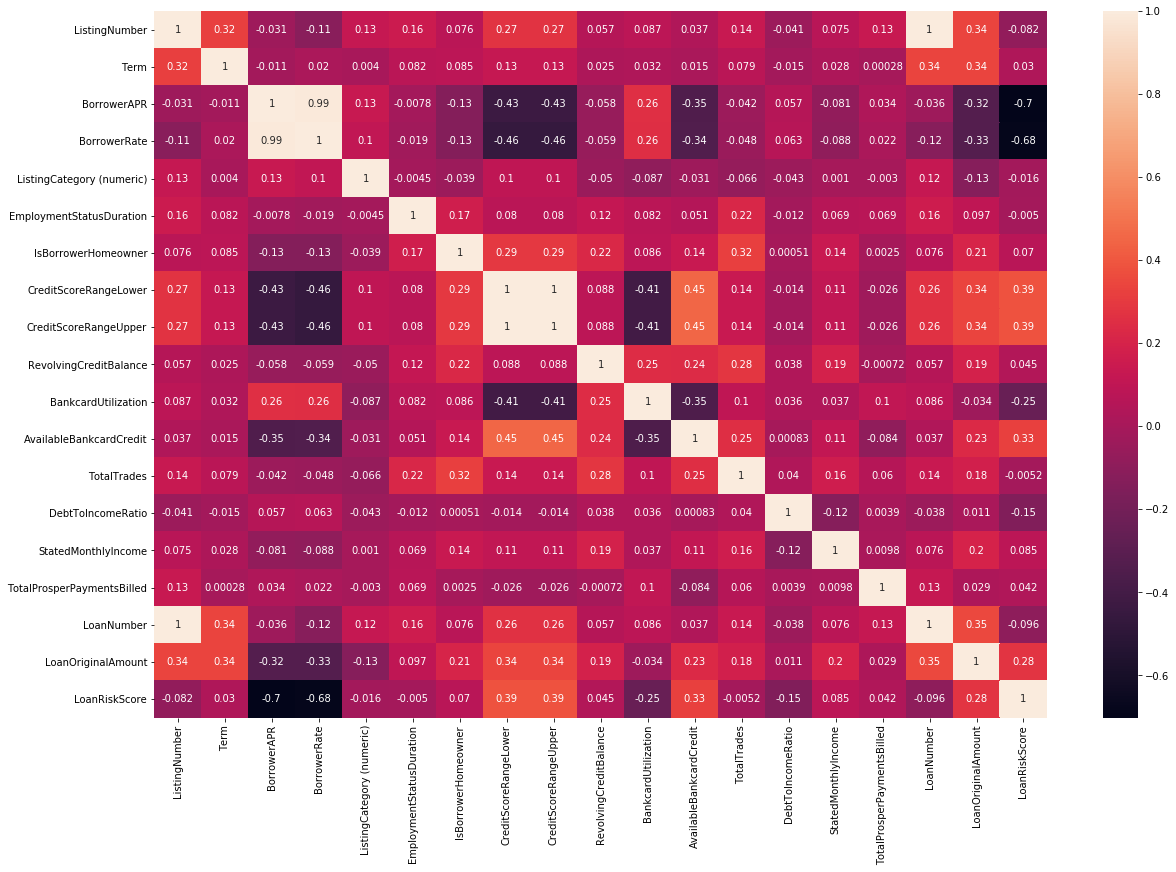
Milestone1 Report

1. **Preprocessing techniques:**

* Drop columns: we remove two columns: (CreditGrade, TotalProsperPaymentsBilled) from the dataset because they have a lot of null values.
* Replace Nonvalues: we replace nonvalues from columns with mean of columns.
* Encode the text features: we convert some features from text to numerical values: (LoanStatus, BorrowerState, EmploymentStatus, IsBorrowerHomeowner, IncomeRange).
* Feature Scaling: we use standard scaler to scale some features: (RevolvingCreditBalance, AvailableBankcardCredit, StatedMonthlyIncome, LoanNumber).

1. **Feature Selection:**
   * we get the features that affect LoanRiskScore by more than 30%
   * The correlation matrix:
2. **Regression techniques:**
   * Model 1:
     + Polynomial Regression
     + MSE testing: 1.6375833151045402
     + MSE training: 1.6339222023696616
     + Training time: 1.6735265254974365
     + Accuracy: 63.89266221570586 %
   * Model 2:
     + Multivariable Regression
     + MSE testing: 2.225830068875934
     + MSE training: 2.2462563222099488
     + Training time: 0.045876264572143555
     + Accuracy: 63.378050758043344 %
3. **The discarded features:**
   * The messing features: (ListingNumber) because it doesn't affect on dataset.
   * The dropped features: (CreditGrade, TotalProsperPaymentsBilled) because they have a lot of nonvalues.
4. **Splitting Dataset:**
   * We are split dataset to 70% training and 30% for testing.
5. **Conclusion:**
   * We first clear data by many several preprocessing techniques to be more accurate to learn with different types of models.
   * We note that polynomial regression takes training time more than linear model and take more time as we raise the degree.
   * The MSE in linear model more than polynomial regression and accuracy is less.