

DataFest 2017

Chapel-Boys.R

4/1/2017

Introduction

We seek to use [OUTSIDE_DATA] to predict the propensity to book. In order, to measure such user behavior we consider a number of classification models and then use a validation procedure to select that one that performs most optimally. Then, we attempt to demonstrate the [OUTSIDE_DATA]’s importance to our model and explanatory power.

PCA

Evidently, the `dest.txt` file contains valuable information in relation to characteristics unique to each destination. However, it is prudent to begin our analysis by shrinking the number of explanatory variables from 144 to a smaller number of principal components.

PC1 & PC2

Hence, we can visualize the first two orthogonal principal components:

