Big Data Case Study in Mobile Analytics

Abstract

<http://adilmoujahid.com/posts/2016/08/interactive-data-visualization-geospatial-d3-dc-leaflet-python/>

In this tutorial, we will use a dataset from a Kaggle competition called ["TalkingData Mobile User Demographics"](https://www.kaggle.com/c/talkingdata-mobile-user-demographics). This dataset is provided by TalkingData, China’s largest third-party mobile data platform. It contains app usage data, geolocation data and mobile device properties.The goal of the competition is to predict the gender and age segments of users based on the data provided.

Data visualization is an important first step in the data analysis workflow. It enables us to effectively discover patterns through graphical means, and to represent these findings in a meaningful and effective way.

The dataset that we will use contains various attributes that can be combined together to build interesting data visualizations. Geospatial data is particularly interesting, as it allows us to see how the user profiles and usage behavior changes based on the location.

In this tutorial, we will build a data visualization that combines a map that shows user locations together with various charts that summarizes users’ information and usage behavior. We will make this visualization interactive, so we can drill down into a particular user segment or location.