Project 2 Report - Databox data visualization

Semir Salkić[1]

Abstract

In this paper databox data insights are presented through text, measurement and statistical visualizations. Two datasets were used which provided data about interactive usage actions taken by companies while using Databox platform. Presented visualizations are taking into consideration actions of specific users while presenting overall statistical inference taken from the data.

Keywords
Databox, Dashboard integration,Statistical Analysis, Data Analysis

[1] *ss9343@student.uni-lj.si, 63190409*

Introduction

This report provides the basic summary and visualiza- tions of the Databox user signups platform data and Databox events data datasets, provided by the company Databox. The first dataset describes how the companies utilize the Databox platform with events, while the second dataset contains details related to each user (i.e. company). Databox platform is cloud-based performance tracking platform for dashboard integration and KPI overview. The dataset for Events consists of 7,948,588 entries and 4 columns: Date, Event, Event Count and Space Id. The second dataset gives detailed information about the specific company (called user) which used the platform. Data is recorded in the time interval from 25^{th} November 2019 to 15^{th} November 2021. campaigns over the course of 2018 and 2019. The second dataset provides useful insight into how each company is defined and what is the structure and action settings of the company profile on the platform. Having this in mind, dataset contains features like country, is agency,timestamps of the company creating and when it became paying etc. This dataset contains 101,699 data points.

Analysis

The Databox events dataset provides information about 34,322 unique users and their interactions with the plaform during period of 2 years (November 2019 - November 2021). The number of the specific event on given day varies from 0 to 422, while mean value is 0.4025. There is 8 distinct events. They are: calculation created, metric created, new connection added, new user added, query created, share new snapshot, share public url by copy/paste, sessions. In Figure 1 event count for sharing public url by copy/paste and for sharing new snapshot is presented. Besides that, every last day of the month is marked in the plot. This is done due to the fact that pattern is emerging during time. We can that there is increase of event counts for each last day of the month. This can be interpreted as work flow of the companies where work on the platform is being shared and backed up by employees, while also creating end of the month report which can be used to update company management of current work progress.

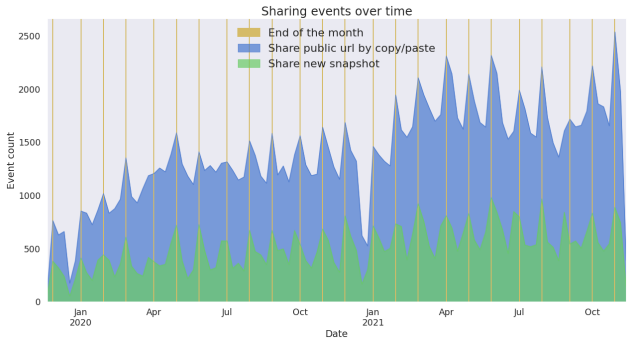

Figure 1. Sharing events over time. Event count is shown during time for sharing events.

In the Figure 2 sessions data is presented categorized with onboarding levels. Besides that, company properties such as if the company is an agency and if the company had a trial.Each bar plot is calculated for 95% C.I. with 10 bootstrap iterations. We can conclude that significant impact on number of session is present on the property of the trial. We can see that all of the categorization of the sessions which used a trial run have larger number of sessions, while information about agency is affecting the sessions in last (3^{rd}) onboarding level.

In the Figure 3 users in space are presented by event categorization. Each company added certain number of users. Company has unique id which is used to calculate added

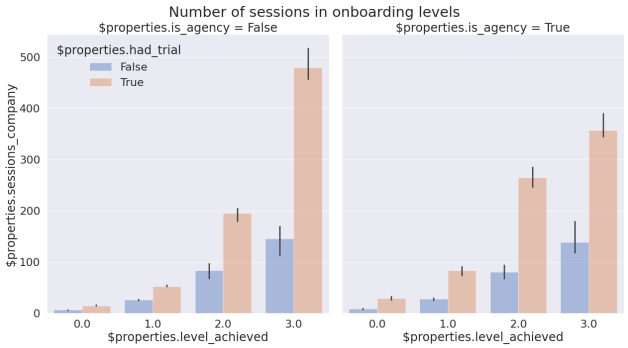

Figure 2. Number of sessions in onboarding levels
Categorized sessions by onboarding level with different categories for companies - If it is an agency and if they had trial of the platform

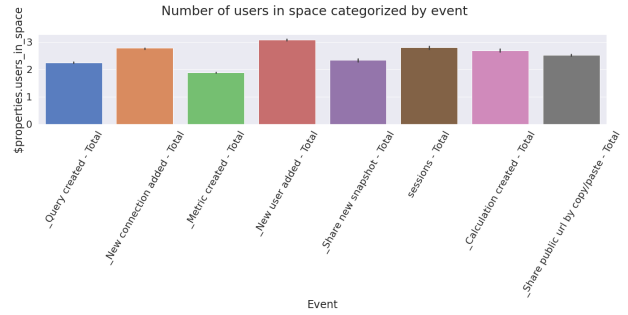

Figure 3. Number of users in space categorized by event
Each unique user is referenced with data from both datasets and presented here

users in the company space. Those users are divided by their interactions with the platform (events).

Conclusion

In this paper observed data insights are presented from the Databox datasets. With provided insight we can conclude that there is certain patterns in the data points which with further processing and transformation can be used to construct data inference and develop user interaction prediction model.