Exploratory Data Analysis

About Datasets

There are 3 sets of images placed in D1, D2 and D3 folders.

D1

|--IMAGES

|--TEXT\_LABELS

Images format- PNG

Annotation format- .txt

# Images

The IMAGES folder contains wireframe (similar) images with different UI components.

| Dataset | Components |
| --- | --- |
| D1 | 'btn-inactive', 'btn-active', 'btn-red', 'btn-orange', 'btn-green' |
| D2 | 'page-title','ck-box','radio','dropdn','slider','search-bar','btn-active','btn-inactive','btn-green','btn-orange','btn-red','icons','cart-link','burger-link','rating','medium-title' |
| D3 | 'field-inactive','img-btn-card','img-txt-btn-card','btn','entry','field-active','category','img-card','user-img','user-img-user-name','logo-txt' |

## D1 Images

There is a pattern in dimension of components. The dimension of UI components is almost same across the dataset. Ex btn-red, btn-orange

There are components which vary only in color but shape is same- Ex btn-red, btn-orange

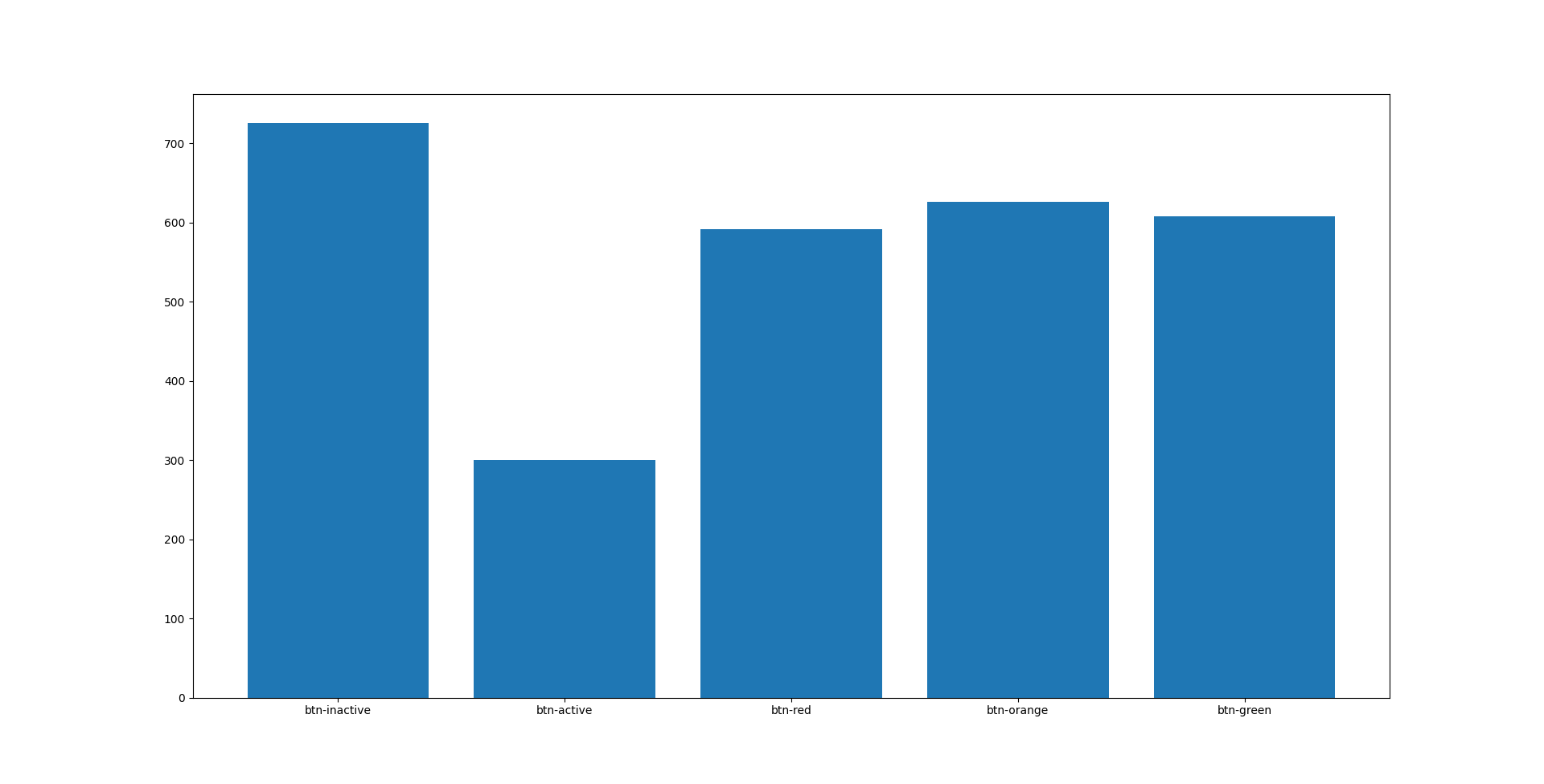
Btn-inactive and btn-active have larger dimension compared to others. The scale of dimension of UI components is same across images.

This dataset is fairly **uniform.** except for btn-inactive.

Max label count: 'btn-inactive'- 726

Min label count: 'slider'- 300

{'btn-inactive': 726, 'btn-active': 300, 'btn-red': 592, 'btn-orange': 626, 'btn-green': 608}



## D2 Images

Max label count: 'btn-inactive'- 748

Min label count: 'slider'- 145

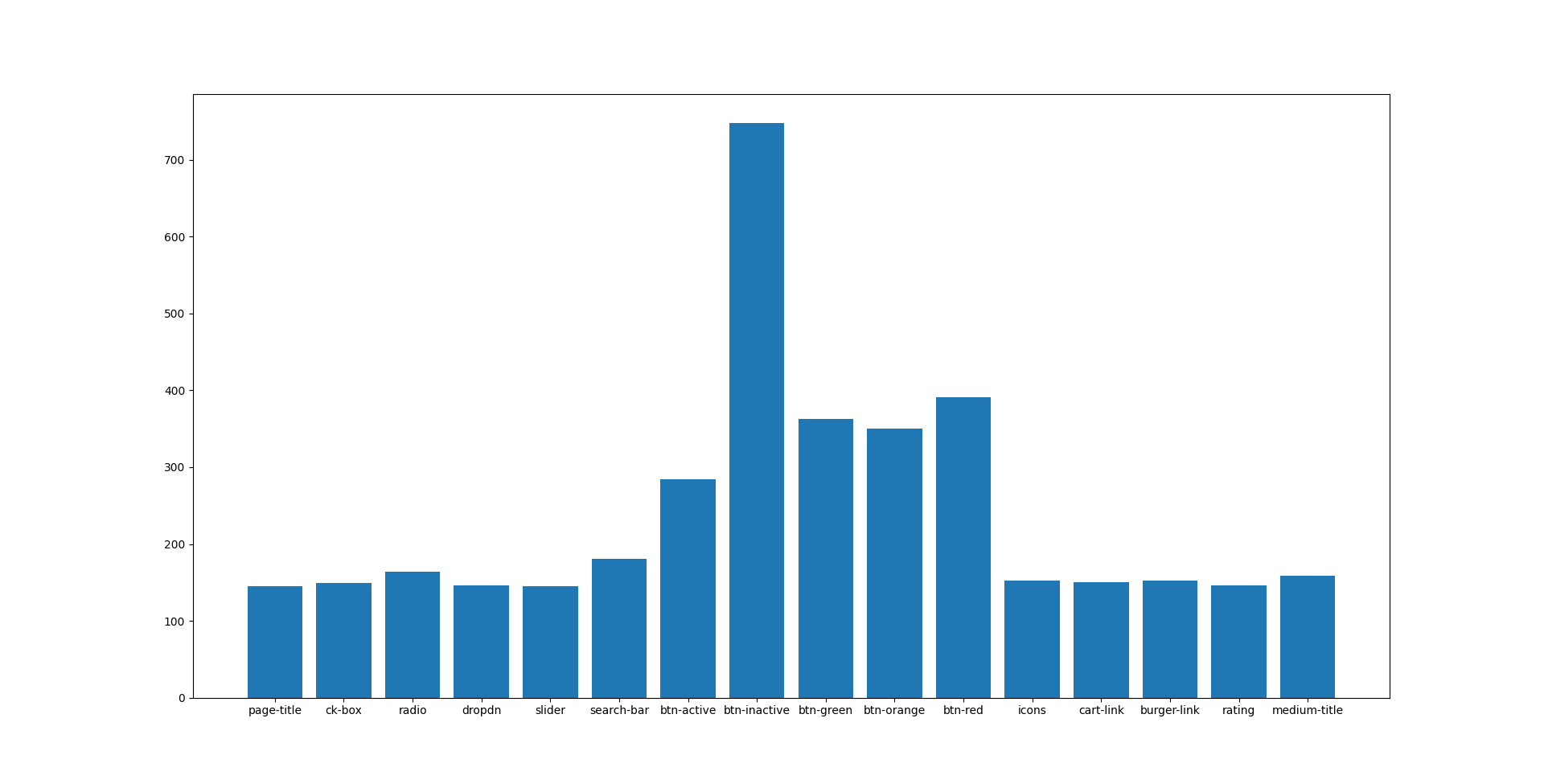
D2 images contains all UI components of D1. In addition they have sidebars, page-title, medium-title.

In this dataset there is huge variation in the components dimensions.

Ex. Slider, Checkbox, icons, cart-link, search-bar all have varied shapes. Rating, checkbox are not continuous too.

In these images **text** is also annotated ex page-title, medium-title

{'page-title': 145, 'ck-box': 149, 'radio': 164, 'dropdn': 146, 'slider': 145, 'search-bar': 181, 'btn-active': 284, 'btn-inactive': 748, 'btn-green': 363, 'btn-orange': 350, 'btn-red': 391, 'icons': 152, 'cart-link': 150, 'burger-link': 152, 'rating': 146, 'medium-title': 159}



## D3 Images

This dataset is completely different from D1 and D3.

Images have home and sidebar panels.

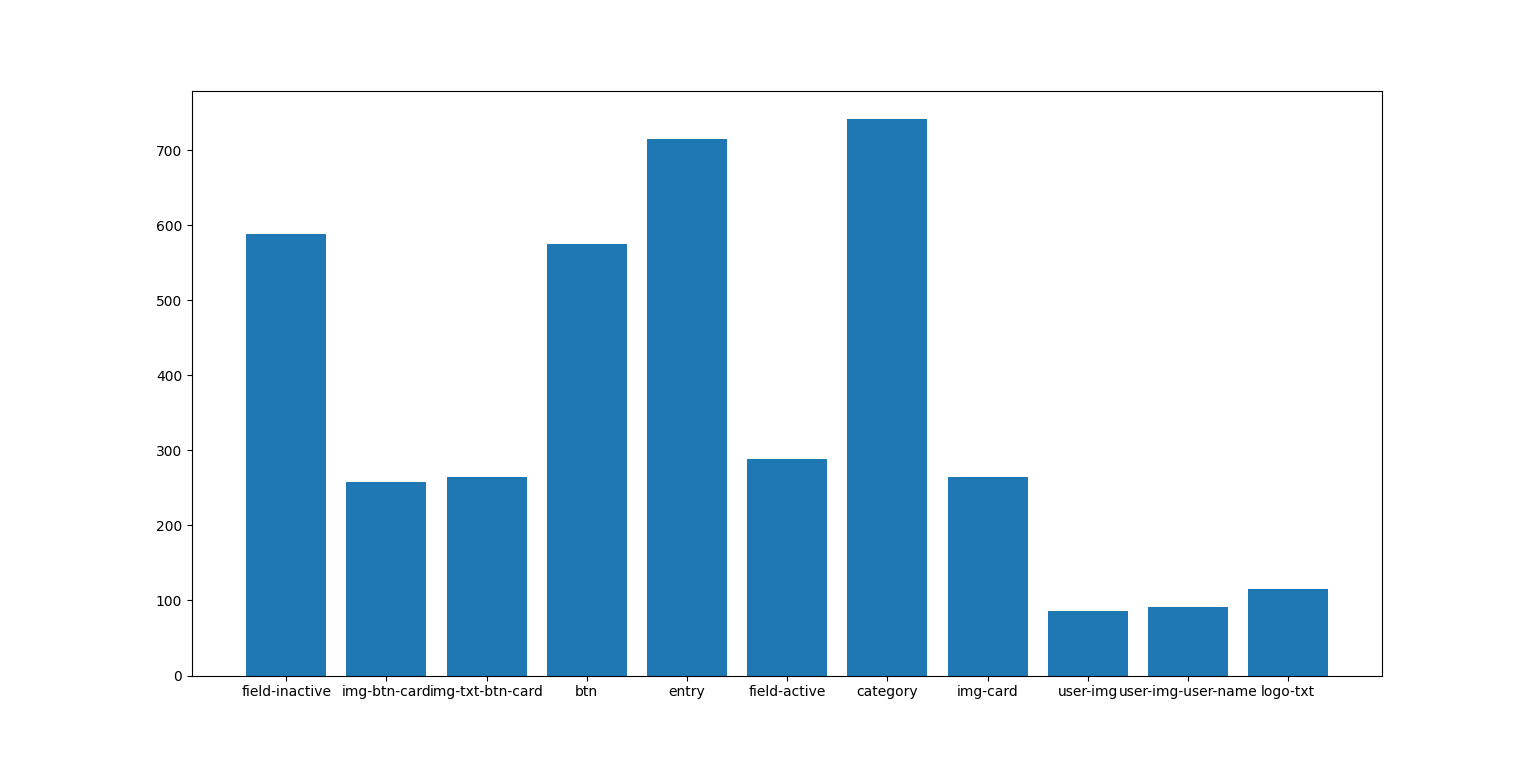
Image/Btn cards are bit similar.

Dataset is **un-balanced**

Max label count: ''user-img'- 86

Min label count: entry- 715

{'field-inactive': 589, 'img-btn-card': 258, 'img-txt-btn-card': 265, 'btn': 575, 'entry': 715, 'field-active': 289, 'category': 742, 'img-card': 265, 'user-img': 86, 'user-img-user-name': 91, 'logo-txt': 115}



# Annotations

The annotations are provided as text files. It cannot be used directly for training any detection model. The data has to be annotated as per YOLO format.