

## Ex Taedio

Remember, a few hours of trial and error can save you several minutes of looking at the README.

— I Am Devloper (@iamdevloper) November 7, 2018

Ex Taedio is a dashboard built using Streamlit and Plotly. Its goal is to help the user create plots and visualizations.

## Getting Started

Ex Taedio is available as an heroku app here but you can also build it yourself following these instructions.

### Prerequisites

You need Python, at least version 3.6, installed on your computer.

### Installing

- Clone the repository at this address: [https://github.com/tr31zh/ask\\_me\\_polotly](https://github.com/tr31zh/ask_me_polotly)
- Move into the created folder
- Create new environment: `python -m venv env`
- Activate environment: `source ./env/bin/activate`
- Clone environment: `pip install -r requirements.txt`
- Run Ex Taedio: `streamlit run sl_plot_me.py`

## Using

- Enable *Show information panels (blue panels with hints and tips)*. if in need of help.
- Advanced functionality is hidden behind the *Advanced mode* checkbox.
- When advanced mode is active, data wrangling, advanced plots and advanced plot settings can be enabled.

### Basic plots

#### Scatter

In a scatter plot, each row of ``data_frame`` is represented by a symbol mark in 2D space.

#### Bar

In a bar plot, each row of ``data_frame`` is represented as a rectangular mark.

## Histogram

In a histogram, rows of ``data_frame`` are grouped together into a rectangular mark to visualize the 1D distribution of an aggregate function ``histfunc`` (e.g. the count or sum) of the value ``y`` (or ``x`` if ``orientation`` is ``h``).

## Violin plot

In a violin plot, rows of ``data_frame`` are grouped together into a curved mark to visualize their distribution.

## Box plot - if you must

In a box plot, rows of ``data_frame`` are grouped together into a box-and-whisker mark to visualize their distribution.

Each box spans from quartile 1 (Q1) to quartile 3 (Q3). The second quartile (Q2) is marked by a line inside the box. By default, the whiskers correspond to the box' edges  $\pm 1.5$  times the interquartile range (IQR:  $Q3 - Q1$ ), see "points" for other options.

## PCA (2D)

**\*\*Principal component analysis (2 dimensions)\*\***

Given a collection of points in two, three, or higher dimensional space, a "best fitting" line can be defined as one that minimizes the average squared distance from a point to the line. The next best-fitting line can be similarly chosen from directions perpendicular to the first. Repeating this process yields an orthogonal basis in which different individual dimensions of the data are uncorrelated. These basis vectors are called principal components, and several related procedures principal component analysis (PCA).

## Advanced plots

### Scatter 3D

In a 3D scatter plot, each row of ``data_frame`` is represented by a symbol mark in 3D space.

### Line

In a 2D line plot, each row of ``data_frame`` is represented as vertex of a polyline mark in 2D space.

### Density heat map

In a density heatmap, rows of ``data_frame`` are grouped together into colored rectangular tiles to visualize the 2D distribution of an aggregate function ``histfunc`` (e.g. the count or sum) of the value ``z``.

### Density contour

In a density contour plot, rows of ``data_frame`` are grouped together into contour marks to visualize the 2D distribution of an aggregate function ``histfunc`` (e.g. the count or sum) of the value ``z``.

### Parallel categories

In a parallel categories (or parallel sets) plot, each row of ``data_frame`` is grouped with other rows that share the same values of ``dimensions`` and then plotted as a polyline mark through a set of parallel axes, one for each of the ``dimensions``.

### Parallel coordinates

In a parallel coordinates plot, each row of ``data_frame`` is represented by a polyline mark which traverses a set of parallel axes, one for each of the ``dimensions``.

### Scatter matrix

Plot a scatter matrix for all selected columns

### PCA (3D)

**\*\*Principal component analysis (3 dimensions)\*\***

Given a collection of points in two, three, or higher dimensional space, a "best fitting" line can be defined as one that minimizes the average squared distance from a point to the line. The next best-fitting line can be similarly chosen from directions perpendicular to the first. Repeating this process yields an orthogonal basis in which different individual dimensions of the data are uncorrelated. These basis vectors are called principal components, and several related procedures principal component analysis (PCA).

### Linear Discriminant Analysis

A generalization of Fisher's linear discriminant, a method used in statistics, pattern recognition, and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.

## Neighborhood Component Analysis

A supervised learning method for classifying multivariate data into distinct classes according to a given distance metric over the data. Functionally, it serves the same purposes as the K-nearest neighbors algorithm, and makes direct use of a related concept termed stochastic nearest neighbors.

## Correlation matrix

Plot correlation matrix

## Deployment

Ex Taedio has been deployed to Heroku. At the moment of writing this readme the deployment can be done with the wizard in heroku's dashboard.

## Built With

- Streamlit - The framework used to build the dashboard.
- Plotly - The plotting library
- Pandas, Numpy - Of course
- Scikit learn - Machine learning

## To do

- Fix trend lines
- Add seaborn version
- Add save restore plot co,figuration

## Contributing

Pull requests are welcome. For major changes, please open an issue first to discuss what you would like to change.

## Authors

- Felicià MAVIANE - tr3lzh

## License

This project is licensed under the MIT License - see the LICENSE file for details.