

Atos for internal use

Assignments

author(s) : Koen Peters

document number : CCT-FFF-XXXX

version :

status : Draft

source : Atos

document date : 6 June 2019­

­

­

number of pages : 5

© Copyright 2019, Atos Nederland B.V. All rights reserved. Reproduction in whole or in part is prohibited without the prior written consent of the copyright owner. For any questions or remarks on this document, please contact Atos, +31622916065.

­­­Contents

[1 Assignment 1 Wine reviews 4](#_Toc10708433)

[2 Assignment 6 Pokemon 5](#_Toc10708434)

List of changes

| version | Date | Description | Author(s) |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

­­­­

# Assignment 1 Wine reviews

Read the wine reviews csv into a jupyter notebook. Do some preliminary Exploratory Data Analysis (inspect first few rows, check datatypes and anything else you deem interesting).

After initial EDA is complete, lets dive deeper to inspect the data using plots.

Use matplotlib or seaborn to create the following visualizations:

(you can peek [here](https://www.kaggle.com/residentmario/univariate-plotting-with-pandas#Univariate-plotting-with-pandas), but they use pandas plotting. Make sure to browse multiple pages!)

- create a bar chart displaying the top 10 most occurring provinces, does any province stand out?

- Enhance the plot by converting the frequency (counts) to percentages, how much does the top 1 province contribute to total wine production?

- create a bar chart of the points awarded to the wines in the review, make sure to have the X-axis in ascending order.

- with 20 categories a bar chart is still easy to interpret, but what if there are 100 categories? Create a line chart of the points awarded to the wines in the review.

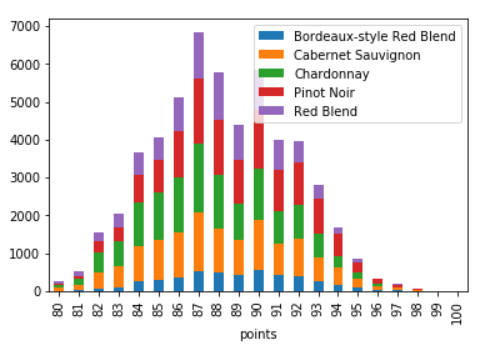
- Enhance the previous plot by filling the area under the curve.

- Histograms are used to view the distribution of a variable, create a histogram of the price per bottle for bottles under $200 dollar.

- Now create a similar histogram including all price ranges, how does this influence your visualization?

- Create a boxplot with wine type on the x-axis and points on the y-axis, what can you say about the spread, quantiles and mean points?

**Challenge**: create a stacked barplot of point rating where you can see how much each wine contributed to the rating (see image)



You can continue exploring the dataset or move on the next assignment.

# Assignment 6 Pokemon

Read the pokemon csv into a jupyter notebook. Do some preliminary Exploratory Data Analysis (inspect first few rows, check datatypes and anything else you deem interesting).

After initial EDA is complete, lets dive deeper to inspect the data using plots.

Use matplotlib or seaborn to create the following visualizations:

(you can peek [here](https://www.kaggle.com/residentmario/univariate-plotting-with-pandas#Univariate-plotting-with-pandas) and [here](https://towardsdatascience.com/exploratory-analysis-of-pokemons-using-r-8600229346fb), but they use different plotting libraries)

- Create a sorted barchart of pokemon type1, which type occurs the most? and which the least?

- Create a frequency plot of pokemon released per generation

- Create a line plot of pokemon HP. Does this seem like a correct visualization? What other type of plot can you make on pokemon HP?

- create a barplot of the legendary pokemon based on their type1, which is the most occuring legendary type? which types are tied?

- Create a scatterplot of attack vs defense.

- Enhance the previous plot by giving legendary pokemon a different collor in the scatterplot than the other pokemon

**Challenge**: create a heatmap/correlation plot of attack type effectiveness, based on type 1 of the pokemon. Which types are effective aggainst dark?

General note: awesome plot examples <https://www.machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-master-plots-python/>