

Portfolio of Data Analytics Projects

Bernardo Di Chiara
dichiarabernardo@gmail.com
+358 45 1330652

This file is extracted from my GitHub profile, which contains my portfolio of data analytics projects.

<https://github.com/BerniHacker/CV/blob/master/README.md>

The projects are divided in the following categories:

- Statistics, Data Cleansing, Data Preparation, Data Visualization
- Prediction Models, Machine Learning
- R Programming
- Python Programming
- Databases

For each file a short description and the link to the relevant file or repository are provided.

Statistics, Data Cleansing, Data Preparation, Data Visualization

File Name	Description	File Link
Time_Use_in_Finland	A Data Cleansing and Visualization Project in Python .	https://github.com/BerniHacker/Time_Use_in_Finland/blob/master/README.md
A_Data_Analysis_Project_in_R.html	This study uses publicly available data to create a research question for which first some exploratory data analysis is done and then inferential statistics is performed to verify if the possible correlations that have been found are statistically relevant. The file has been produced in R by using RStudio.	http://htmlpreview.github.io/?https://github.com/BerniHacker/CV/blob/master/A_Data_Analysis_Project_in_R.html
Getting and Cleaning Data Project	This is a link to a repository containing a script in R to upload, clean and prepare some data sets. A Code Book is also contained.	https://github.com/BerniHacker/GettingAndCleaningData
A_Plotting_Project_in_Python_using_Matplotlib.html	This project consists in data manipulation and plotting composite graphs by using Matplotlib .	http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/A_Plotting_Project_in_Python_using_Matplotlib.html
Creating_a_Customized_Visualization.html	This project is a Python implementation of an idea discussed in a paper from Ferreira, Fisher and Konig, which allows visualizing a plot that helps in making judgements about probabilistic data generated through samples. It produces an interactive plot. (In order to test the interactive feature, copy and paste the source code to your Jupyter Notebook environment.)	http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Creating_a_Customized_Visualization.html
Hypothesis_Test_with_Python_Pandas.html	This project consists in retrieving and manipulating data from different files and then performing a hypothesis test in Python .	http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Hypothesis_Test_with_Python_Pandas.html
Data_Retrieval_Cleansing_Manipulation_with_Pandas.html	In this project, three files are loaded in Python as Pandas dataframes. Those dataframes are cleansed and then merged into a single data set. The dataset is further manipulated to find average values, max values, etc ... New columns are added to calculate new variables. Then data from a Python dictionary is added to do further data manipulation and new dataframes are created. The work has been done by using Jupyter Notebook.	http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Data_Retrieval_Cleansing_Manipulation_with_Pandas.html

Prediction Models, Machine Learning

File Name	Description	File Link
Fitting_a_Multiple_Linear_Regression_Model.html	In this document a data set is analyzed, a relevant research question is created and some exploratory data analysis is performed in Python . Then a linear regression model is developed and this model is used to do some prediction.	http://htmlpreview.github.io/?https://github.com/BerniHacker/CV/blob/master/Fitting_a_Multiple_Linear_Regression_Model.html
A_Prediction_Model_Project_in_R.html	This project consists in building and comparing different prediction models in R .	http://htmlpreview.github.io/?https://github.com/BerniHacker/CV/blob/master/A_Prediction_Model_Project_in_R.html

Python Programming

File Name	Description	File Link
Finding10MostCommonWords.py	A simple Python script using dictionaries, lists and tuples	https://github.com/BerniHacker/Python/blob/master/Finding10MostCommonWords.py
MostFrequentEmailAddress.py	A simple Python script using dictionaries	https://github.com/BerniHacker/Python/blob/master/MostFrequentEmailAddress.py
ComputePay.py	A simple Python script using a function	https://github.com/BerniHacker/Python/blob/master/CalculatePay.py
MaxAndMinNumber.py	A simple Python script using conditional statements and loops	https://github.com/BerniHacker/Python/blob/master/MaxAndMinNumber.py
WorkingWithFiles.py	A simple Python script using a file	https://github.com/BerniHacker/Python/blob/master/WorkingWithFiles.py
WorkingWithLists.py	A simple Python script using lists	https://github.com/BerniHacker/Python/blob/master/WorkingWithLists.py

R Programming

File Name	Description	File Link
A_Programming_Project_in_R.html	This project consists in creating a function and gives an example of my programming skills in R.	http://htmlpreview.github.io/?https://github.com/BerniHacker/CV/blob/master/A_Programming_Project_in_R.html

Databases

File Name	Description	File Link
Profiling_and_Analyzing_in_SQL.sql	This project consists in profiling and analyzing a dataset in SQL .	https://github.com/BerniHacker/SQL/blob/master/Profiling_and_Analyzing_in_SQL.sql
A_Spark_Job_in_Python.txt	Joining data with Spark by using Python code and Cloudera VM.	https://github.com/BerniHacker/CV/blob/master/A_Spark_Job_in_Python.txt
A_MapReduce_Job_in_Python.txt	Joining data with MapReduce by using streaming with Python code and Cloudera VM.	https://github.com/BerniHacker/CV/blob/master/A_MapReduce_Job_in_Python.txt