

# Portfolio of Data Analytics Projects

Bernardo Di Chiara  
dichiarabernardo@gmail.com

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:

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

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

Data Dumping		
File Name	Description	File Link
dump_SQL_table.sh	A template for dumping a table from one MySQL database to another MySQL database. The copy is done in chunks and therefore the script can be used also for huge tables.	<a href="https://github.com/BerniHacker/Linux/blob/master/dump_SQL_table.sh">https://github.com/BerniHacker/Linux/blob/master/dump_SQL_table.sh</a>
SQL_to_CSV.py	This script queries the full content of a defined table of a defined MySQL database (with defined credentials) and stores the result of the query temporarily into a pandas dataframe. The content of the dataframe is then dumped into a file with a defined name. The script allows handling "large data". Both the SQL query and the dump into the CSV file are performed in chunks according to pre-defined parameter values.	<a href="https://github.com/BerniHacker/Python/blob/master/SQL_to_CSV.py">https://github.com/BerniHacker/Python/blob/master/SQL_to_CSV.py</a>

Statistics, Data Cleansing, Data Preparation, Data Visualization		
File Name	Description	File Link
Time_Use_in_Finland	A Data Cleansing and Visualization Project in <b>Python</b> .	<a href="https://github.com/BerniHacker/Time_Use_in_Finland/blob/master/README.md">https://github.com/BerniHacker/Time_Use_in_Finland/blob/master/README.md</a>
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 <b>R</b> by using RStudio.	<a href="http://htmlpreview.github.io/?https://github.com/BerniHacker/CV/blob/master/A_Data_Analysis_Project_in_R.html">http://htmlpreview.github.io/?https://github.com/BerniHacker/CV/blob/master/A_Data_Analysis_Project_in_R.html</a>
Getting and Cleaning Data Project	This is a link to a repository containing a script in <b>R</b> to upload, clean and prepare some data sets. A Code Book is also contained.	<a href="https://github.com/BerniHacker/GettingAndCleaningData">https://github.com/BerniHacker/GettingAndCleaningData</a>
A_Plotting_Project_in_Python_using_Matplotlib.html	This project consists in data manipulation and plotting composite graphs by using <b>Matplotlib</b> .	<a href="http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/A_Plotting_Project_in_Python_using_Matplotlib.html">http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/A_Plotting_Project_in_Python_using_Matplotlib.html</a>
Creating_a_Customized_Visualization.html	This project is a <b>Python</b> 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.)	<a href="http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Creating_a_Customized_Visualization.html">http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Creating_a_Customized_Visualization.html</a>
Hypothesis_Test_with_Python_Pandas.html	This project consists in retrieving and manipulating data from different files and then performing a hypothesis test in <b>Python</b> .	<a href="http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Hypothesis_Test_with_Python_Pandas.html">http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Hypothesis_Test_with_Python_Pandas.html</a>
Data_Retrieval_Cleansing_Manipulation_with_Pandas.html	In this project, three files are loaded in Python as <b>Pandas</b> 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.	<a href="http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Data_Retrieval_Cleansing_Manipulation_with_Pandas.html">http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Data_Retrieval_Cleansing_Manipulation_with_Pandas.html</a>

## 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 <b>Python</b> . Then a linear regression model is developed and this model is used to do some prediction.	<a href="http://htmlpreview.github.io/?https://github.com/BerniHacker/CV/blob/master/Fitting_a_Multiple_Linear_Regression_Model.html">http://htmlpreview.github.io/?https://github.com/BerniHacker/CV/blob/master/Fitting_a_Multiple_Linear_Regression_Model.html</a>
A_Prediction_Model_Project_in_R.html	This project consists in building and comparing different prediction models in <b>R</b> .	<a href="http://htmlpreview.github.io/?https://github.com/BerniHacker/CV/blob/master/A_Prediction_Model_Project_in_R.html">http://htmlpreview.github.io/?https://github.com/BerniHacker/CV/blob/master/A_Prediction_Model_Project_in_R.html</a>

## Python Programming (continues)

File Name	Description	File Link
Scraping_Web_Pages.py	The program asks the user to insert a URL, reads the HTML, searches for a link that is in a user-provided position respect to the top of the page and follows that link. The process is repeated a user-defined number of times and returns the visible text corresponding to the last found link. Validation of user input is performed.	<a href="https://github.com/BerniHacker/Python/blob/master/Scraping_Web_Pages.py">https://github.com/BerniHacker/Python/blob/master/Scraping_Web_Pages.py</a>
Using_a_Web_Service.html	This program prompts for a location, contacts a web service, retrieves JSON data corresponding to the web service and the user-defined location and finds some data.	<a href="http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Using_a_Web_Service.html">http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Using_a_Web_Service.html</a>
Extracting_JSON_Data.html	This program prompts for a URL, reads the JSON data from that URL, extracts some data and does some calculation.	<a href="http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Extracting_JSON_Data.html">http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Extracting_JSON_Data.html</a>
Extracting_XML_Data.py	This program prompts for a URL, reads the XML data from that URL, extracts some data and does some calculation.	<a href="https://github.com/BerniHacker/Python/blob/master/Extracting_XML_Data.py">https://github.com/BerniHacker/Python/blob/master/Extracting_XML_Data.py</a>
Finding10MostCommonWords.py	A simple Python script using dictionaries, lists and tuples	<a href="https://github.com/BerniHacker/Python/blob/master/Finding10MostCommonWords.py">https://github.com/BerniHacker/Python/blob/master/Finding10MostCommonWords.py</a>
MostFrequentEmailAddress.py	A simple Python script using dictionaries	<a href="https://github.com/BerniHacker/Python/blob/master/MostFrequentEmailAddress.py">https://github.com/BerniHacker/Python/blob/master/MostFrequentEmailAddress.py</a>

## Python Programming (follows)

File Name	Description	File Link
ComputePay.py	A simple Python script using a function	<a href="https://github.com/BerniHacker/Python/blob/master/CalculatePay.py">https://github.com/BerniHacker/Python/blob/master/CalculatePay.py</a>
MaxAndMinNumber.py	A simple Python script using conditional statements and loops	<a href="https://github.com/BerniHacker/Python/blob/master/MaxAndMinNumber.py">https://github.com/BerniHacker/Python/blob/master/MaxAndMinNumber.py</a>
WorkingWithFiles.py	A simple Python script using a file	<a href="https://github.com/BerniHacker/Python/blob/master/WorkingWithFiles.py">https://github.com/BerniHacker/Python/blob/master/WorkingWithFiles.py</a>
WorkingWithLists.py	A simple Python script using lists	<a href="https://github.com/BerniHacker/Python/blob/master/WorkingWithLists.py">https://github.com/BerniHacker/Python/blob/master/WorkingWithLists.py</a>

## Databases

File Name	Description	File Link
Profiling_and_Analyzing_in_SQL.sql	This project consists in profiling and analyzing a dataset in <b>SQL</b> .	<a href="https://github.com/BerniHacker/SQL/blob/master/Profiling_and_Analyzing_in_SQL.sql">https://github.com/BerniHacker/SQL/blob/master/Profiling_and_Analyzing_in_SQL.sql</a>
A_Spark_Job_in_Python.txt	Joining data with <b>Spark</b> by using Python code and Cloudera VM.	<a href="https://github.com/BerniHacker/CV/blob/master/A_Spark_Job_in_Python.txt">https://github.com/BerniHacker/CV/blob/master/A_Spark_Job_in_Python.txt</a>
A_MapReduce_Job_in_Python.txt	Joining data with <b>MapReduce</b> by using streaming with Python code and Cloudera VM.	<a href="https://github.com/BerniHacker/CV/blob/master/A_MapReduce_Job_in_Python.txt">https://github.com/BerniHacker/CV/blob/master/A_MapReduce_Job_in_Python.txt</a>
Working_with_Multiple_SQL_Tables_With_Python.html	This <b>Python</b> code creates a SQL database with multiple tables, retrieves a JSON file, extracts data from the file and uses it to populate the database.	<a href="http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Working_with_Multiple_SQL_Tables_With_Python.html">http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Working_with_Multiple_SQL_Tables_With_Python.html</a>
Creating_Modifying_and_Querying_a_SQL_Table_with_Python.html	This <b>Python</b> application reads the mailbox data contained in a text file, counts the number of email messages per organization (i.e. domain name of the email address) and stores the result in a SQLite database.	<a href="http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Creating_Modifying_and_Querying_a_SQL_Table_with_Python.html">http://htmlpreview.github.io/?https://github.com/BerniHacker/Python/blob/master/Creating_Modifying_and_Querying_a_SQL_Table_with_Python.html</a>

## 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.	<a href="http://htmlpreview.github.io/?https://github.com/BerniHacker/CV/blob/master/A_Programming_Project_in_R.html">http://htmlpreview.github.io/?https://github.com/BerniHacker/CV/blob/master/A_Programming_Project_in_R.html</a>