## Tools for Data Analytics

## Student Submission Form

To complete this assessment, you need to create a ZIP archive folder that includes all the files and code used complete the data analysis organized in three subfolders named “Part A”, “Park B,” and “Part C,” as they related to the part of the ask, and a completed copy of this form in the main folder. You will upload your zipped folder that includes this completed form and the subfolders to Taskstream to complete your submission. Use as many rows in the tables below as necessary and remove those not used.

**Student Name: Chase Clark**

**WGU Student ID: 000361224**

**Inventory of Part A-related files (Subfolder “Part A”):**

**Version of Python used:2.7**

**Python Libraries used (if any):requests,bs4, re, csv,os**

**Platform Python was used on:windows 10, cmdline**

**Name of the PDF file with the responses to task prompts from Part A: C742 Project - Final**

**Name(s) and description(s) of Python files:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Extension** | **Description** |
| **scraper** | **.py** | **Webscraper** |
| **\_** | **\_** | **\_** |

**Name(s) and description(s) of input file(s) from Part A:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Extension** | **Description** |
| **\_** | **\_** | **\_** |
| **\_** | **\_** | **\_** |

**Name(s) and description(s) of output file(s) from Part A:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Extension** | **Description** |
| **links** | **csv** | **Output for all links** |
| **htmlFile** | **.txt** | **HTML output** |

**Inventory of Part B-related files (Subfolder “Part B”):**

**SQL Environment used: MySQL Workbench**

**Platform SQL was used on: Windows 10**

**Name of the PDF file with the responses to task prompts from Part B: C742 Project - Final**

**Name(s) and description(s) of SQL code files:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Extension** | **Description** |
| **queriesFinal** | **sql** | **Query used in part J** |
| **\_** | **\_** | **\_** |

**Name(s) and description(s) of input file(s) from Part B:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Extension** | **Description** |
| **sqlInsert** | **.xlsx** | **This file contains the populations values as well as the excel formula used to generate the insert statements** |
| **\_** | **\_** | **\_** |

**Name(s) and description(s) of output file(s) from Part B:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Extension** | **Description** |
| **popDifferencesPartJ** | **csv** | **Output for part J** |
| **\_** | **\_** | **\_** |

**Inventory of Part C-related files (Subfolder “Part C”):**

**R version used: 3.4.1**

**R packages used (if any): \_**

**Platform R was used on: windows 10**

**Name of the PDF file with the responses to task prompts from Part C: C742 Project - Final**

**Name(s) and description(s) of R script(s):**

|  |  |  |
| --- | --- | --- |
| **Name** | **Extension** | **Description** |
| **2020 prediction** | **R** | **Creates linear regression model for dataset** |
| **2020histo** | **R** | **Creates histogram for data set** |

**Name(s) and description(s) of input file(s) from Part B:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Extension** | **Description** |
| floPopEst | **tab** | **Contains census population values and estimated population values** |
| **\_** | **\_** | **\_** |