THE UNIVERSITY OF BIRMINGHAM

UNIVERSITY OF
BIRMINGHAM

Bootcamp Data analytics
Pandas-Challenge

Date of report/submission: 2nd of Feb. 2020
Lecturer: Dr Antony

by Alemseghed Ghebrezghi
EMAIL: AT.GHEBREZGHI@GMAIL.COM

CONTENTS

1. Introduction
One of the aim of data analysis is to have insights of school perromances based on aggregate information which quantitative measuring values of statistics using pandas and visualisation of the performance, comparison of different schools on maths and reading.

2. Analysis
From the given raw data various insights have been found. These insights have been presented using different means such as descriptive statistics. These are values have been found using pandas.
This information which has been found includes.
- District summary of various schools which includes.
  - Total number of unique schools
  - Total students
  - Total budget
  - Average math score
  - Average reading score
  - % Passing math (the percentage of students who passed math)
  - % Passing reading (the percentage of students who passed reading)
  - % Overall passing (the percentage of students who passed math and reading)
- School summary

These include:
  - School name
  - School type
  - Total students
  - Total school budget
  - Per student budget
  - Average math score
  - Average reading score
  - % Passing math (the percentage of students who passed math)
  - % Passing reading (the percentage of students who passed reading)
  - % Overall passing (the percentage of students who passed math AND reading)

- Highest-Performing Schools (by % Overall Passing)
- Lowest-Performing Schools (by % Overall Passing)
- Math Scores by Grade
- Reading Scores by Grade
- Scores by School Spending
- Scores by School Size
- Scores by School Type

These all have been found using python and pandas application. In doing so different types of analysis have been carried out. Such as importing of CSV file, loading, and converting into DataFrames, merging of DataFrames, grouping, inserting and slicing and so on.

Conclusion
In summary, application of pandas helps a lot in deriving insights from raw data. These information incudes top statistical values, comparisons, categorisation by types, by sizes and so on.
However, it would have been helpful to identify and work on missing data and interpolation and extrapolation could have been done to estimates the future drawbacks of the schools efficiency.