**Creation Documentation: Dental and Student Enrolment in Northland**

1. **Purpose**

This document outlines the creation of the NDHB\_School\_data file create as part of the Northland GovHack, 2019 event. This includes:

* Reviewing the data used,
* Describing the process for manipulating the data, and
* notes about the final product, uses, and limitations.

The R script used is attached in Appendix A.

1. **School and Dental Enrolment Data**

There were three data sources used for this portion of the project:

1. Counts of Students Enrolled in dental programs by school in Northland,
2. School Directory information provided through Education Counts API, and
3. Counts of School Enrolment by age, gender, ethnicity from Education Counts download.

**II.1 Students Enrolled**

Provided by request directly through the Northland DHB in September, 2019; representing the students in the system current at the time of request. The data contains the counts of students known to be enrolled in the CDA scheme. This is recorded once the student has been approved and recorded with a dentist. This was split into 3 tables, the total count of student by school between the ages of:

* 4 and 19
* 13 and 19, and
* 4 and 14

**II.2 School Directory**

The school directory data was downloaded from the Education Counts Data Service using the URL link provided:

[Page Link: https://catalogue.data.govt.nz/dataset/directory-of-educational-institutions/resource/bdfe0e4c-1554-4701-a8fe-ba1c8e0cc2ce](Page%20Link:%20https://catalogue.data.govt.nz/dataset/directory-of-educational-institutions/resource/bdfe0e4c-1554-4701-a8fe-ba1c8e0cc2ce)

Table URL: <https://catalogue.data.govt.nz/dataset/2756db90-a096-4ffa-9fb4-73b74ad279d9/resource/bdfe0e4c-1554-4701-a8fe-ba1c8e0cc2ce/download/schooldirectory-07-09-2019-233057.csv>

The data generally contains descriptive information about all schools in New Zealand including Name, deciles, contacts, location (address and coordinates), and summary information about total enrolments. The data was extracted on September 7, 2019.

**II.3 School Enrolment**

School enrolment data was retrieved as a zipped csv filed from the Education Counts Statistics page under Schooling:

Page: <https://www.educationcounts.govt.nz/statistics/schooling/student-numbers/6028>

File: <https://www.educationcounts.govt.nz/__data/assets/file/0005/195359/Machine-Readable-Roll-by-Age-and-Ethnicity-2009-2018.zip>

The data is in a long form which lists the number of students by school and year which are in each age, gender, and ethnic grouping. The data was extracted on September 7, 2019 and covered the period of 2008 to 2018 (capturing July 1 counts).

1. **Data Processing in R.**

Data was imported and manipulated in R version 3.5.3. After importing the data, there were \*\*\* steps in combining the 3 different sources of data.

The first step was to merge the Dental Enrolment fields to the School Directory table. Second, the School enrolment data had to be summarized by school, to capture the counts that corresponded with the age groupings provided in the Dental Enrolment data. Once summarized, the count of student by broad grouping were merged as new fields in the School Directory table.   
  
With the three data sources merged, for each age grouping 2 new fields were calculated:

* Rate- ratio of children enrolled with a dentist to total students
* Difference- total difference between students enrolled and total students

Details on the process can be seen in Appendix A.

1. **Use and Limitations**

This product should be considered in a draft or prototype state, and is not shared in any official capacity. The creators of this data and subsequent media continent are not responsible for how the information is used or shared, and have no liability for any potential errors or inaccuracies.

Users should note the limitations to the data used and information presented. These include:

1. Data incongruence- While the Dental Enrolment and School Directory data are both representative of near midyear 2019 counts, the School Enrolment data is from mid 2018. Because of this discrepancy, there are likely errors in the rates and counts shown.
   1. This issue lead to smaller schools being omitted from the final data table, as their total dental enrolment was larger than the total of student for the age group.
2. Matching errors- Not all schools had a direct match between the School name field from the Dental Enrolment data and the Education Counts tables. This is likely due to clerical errors or difference in how the school’s name is recorded. While a list of schools without a match was generated, they have not been vetted or corrected.
3. Quality and Completeness- Given the time constraints, the data sourced and generated for this project was not vetted for errors or completeness. This may results in schools being omitted or inaccurate information being presented.

**Appendix A.**

library("reshape2", lib.loc="~/R/win-library/3.5")

library("tidyr", lib.loc="~/R/win-library/3.5")

library("xlsx", lib.loc="~/R/win-library/3.5")

library("stringr", lib.loc="~/R/win-library/3.5")

library("dbplyr", lib.loc="~/R/win-library/3.5")

library("tidyverse", lib.loc="~/R/win-library/3.5")

#Reading csv from education counts API which contains all school information and enrollment summary data

EC\_data\_All <- read.csv("https://catalogue.data.govt.nz/dataset/2756db90-a096-4ffa-9fb4-73b74ad279d9/resource/bdfe0e4c-1554-4701-a8fe-ba1c8e0cc2ce/download/schooldirectory-06-09-2019-233055.csv")

#reading csv downloaded from Gdocs containing school enrolment data provided as an extract by Northland DHB, csv separated by number of students enrolled by specific age ranges

DHB\_dataAll <- read.csv("C:\\...GovHack\\DHB\_Teeth\_All.csv")

DHB\_dataU14 <- read.csv("C:\\...\\DHB\_Teeth\_U14.csv")

DHB\_data1418 <- read.csv("C:\\...\\DHB\_Teeth14\_18.csv")

#merging education counts school data with 3 age grouping dfs to add dental enrollment counts

Teeth\_data <- merge(EC\_data\_All, DHB\_dataAll, by.x = 'Org\_Name', by.y = 'School', all.x=T)

Teeth\_data <- merge(Teeth\_data, DHB\_dataU14, by.x = 'Org\_Name', by.y = 'School', all.x=T )

Teeth\_data <- merge(Teeth\_data, DHB\_data1418, by.x = 'Org\_Name', by.y = 'School', all.x=T)

######Creating counts of students by age groupings using machine readable csv from Education Counts, subsetting for the year 2018 and Northland schools removing, charters and punctuations from age field, converting age to numeric

EC\_students <- read.csv("C\\...\\Machine Readable-Roll by Age and Ethnicity 2009-2018.csv") %>% subset (Year..As.at.1.July == '2018' & Region..Regional.Council == 'Northland Region')

EC\_students$Student..Age <- str\_remove\_all(as.character(EC\_students$Student..Age), "Age ") %>% gsub("[[:punct:]]","",.) %>% as.numeric(.)

#Counts for all students by school

Count\_All <- EC\_students %>% subset(Student..Age >4 & Student..Age <19) %>% group\_by(School..Name) %>% summarise(StudAll = sum(ï..Students..â...Values.))

#Counts for 14 to 18 students by school

Count\_1418<- EC\_students %>% subset(Student..Age >13 & Student..Age <19) %>% group\_by(School..Name) %>% summarise(Stud1418 = sum(ï..Students..â...Values.))

#counts for U14 by school

Count\_U14 <- EC\_students %>% subset(Student..Age > 4 & Student..Age < 14 ) %>% group\_by(School..Name) %>% summarise(StudU14 = sum(ï..Students..â...Values.))

#Merging Field counts into final df

Final\_Data <- merge(Teeth\_data,Count\_All, by.x = 'Org\_Name', by.y = 'School..Name', all.y = T)

Final\_Data <- merge(Final\_Data, Count\_U14, by.x = 'Org\_Name', by.y = 'School..Name', all.x = T)

Final\_Data <- merge(Final\_Data, Count\_1418, by.x = 'Org\_Name', by.y = 'School..Name', all.x = T)

#creating and calculating new rate and count fields, replacing NA values with Zeros, filtering for subset which removes schools with impossible rates (over 100% enrolled)

Final\_Data <- Final\_Data %>% mutate(All\_rate = ChildrenAll/StudAll, U14\_rate = ChildrenU14/StudU14, S1418\_rate = Children1418/Stud1418, All\_Diff = StudAll - ChildrenAll, U14\_Diff = StudU14 - ChildrenU14, S1418\_Diff = Stud1418 - Children1418)

Final\_Data[is.na(Final\_Data)] <- 0

Final\_Export <- subset(Final\_Data,All\_rate < 1 & U14\_rate <1 & S1418\_rate < 1)

#writing subset to csv and xlsx files

write.csv2(Final\_Export, "C:\\...\\NDHB\_School\_DentaEnrl2.csv")

write.xlsx(Final\_Export,"C:\\...\\NDHB\_School\_DentaEnrl2.xlsx")