NHS Dentist Statistic for England – 2018-19

1. Introduction

NHS Digital regularly publish data sources on NHS website and provide key insights about dental activities in the England by publishing reports. These data sources are available for public use and can be further explored to uncover new insights. This report aims to utilize NHS dental patient and dentist data to establish the key finding such as relationship between number of patients and their age group, merging IMD deprivation data with dental data and explore the effect of deprivation on dental setup and patients.

The objective of this report is to perform data characterisation, data profiling , data quality checks and create an integrated dataset to produce visualization and detailed analysis to answer key questions.

Data sources for this report are:

NHS Dental Data 2018-19

Context	File Name
Patient Seen File 1	nhs-dent-stat-eng-jan-jun-19-anx3-ps-prac.csv
Patient Seen File 2	nhs-dent-stat-eng-jul-dec-18-anx3-ps-prac.csv
Dentist workforce	NHS Dental Statistics for England 2018-19 Annex3_Workforce.csv

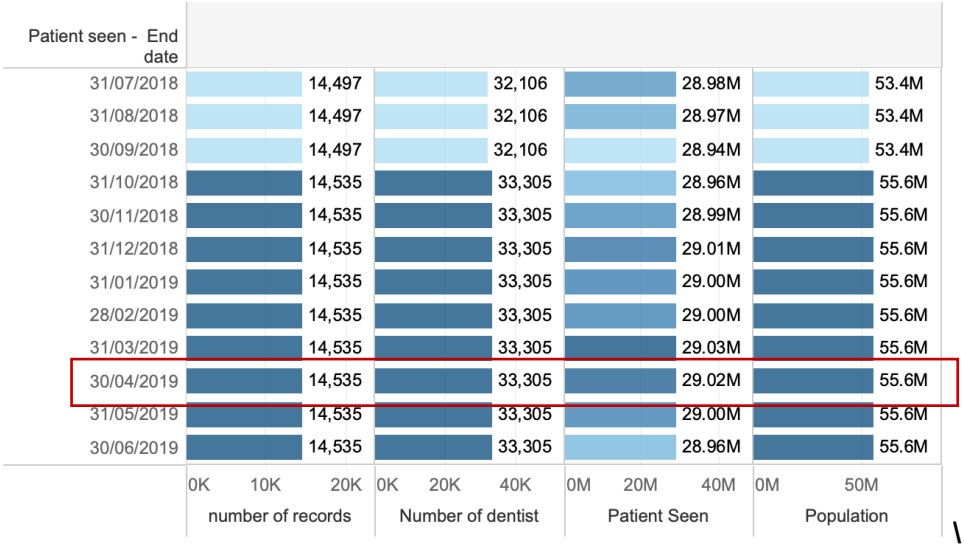
Indices of deprivation
File_13_-_IoD2019_Clinical_Commissioning_Group__CCG__Summaries.xlsx

An integrated dataset is created for 12 months to 31/03/2019 as per given in NHS Dentist Website. However “patient seen end date” in this dataset is from Jun 2018 to Sep 2019 for this data period.

As illustrated by the trend across “patient seen end date” in Fig 1.1, it seems that the data is consistent in volume and there is no noticeable variation in population and other variables.

Hence, further analysis is carried out by sampling the “patient end date” as April 2019.

Fig 1.1 Quick analysis of the patient seen end date data volume



2. Data characterisation

Data characterisation is the summarization of general characteristics of the data which involves below activities in detail:

a. Classification and datatype of the variable – dataset contains 10 variables with datatype as object, int and float

Table 2.1: Dataset variable data type and description

ID	Variable	Variable Description	Variable Classification	Data type
1	CCG_CODE	NHS England Clinical Commissioning Group (CCG) Code	Nominal	object
2	CCG_ONS_CODE	NHS England Clinical Commissioning Group (CCG) ONS Code	Nominal	object
3	CCG_NAME	NHS England Clinical Commissioning Group (CCG) Name	Nominal	object
4	PATIENT_TYPE	Type of patient - Adult or Child	Nominal	object
5	AGE_BAND	Age group of patient	Nominal	object
6	PATIENTS_SEEN	The number of patients seen in the period	Numerical	int64
7	POPULATION	The population estimate in the period	Numerical	int64
8	DENTIST_AGE_GROUP	Age group of dentist	Nominal	object
9	NUMBER_OF_DENTIST	Number of dentist	Numerical	int64
10	AVG_IMD_SCORE	Average rating of deprivation	Numerical	float64

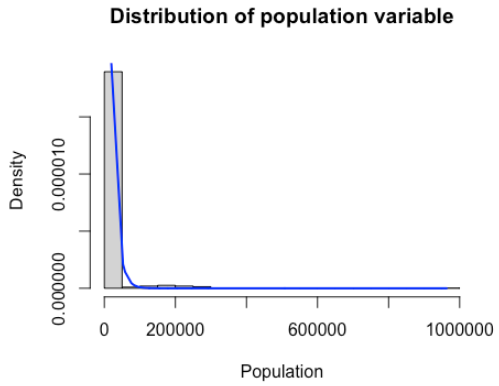
b. Data distribution

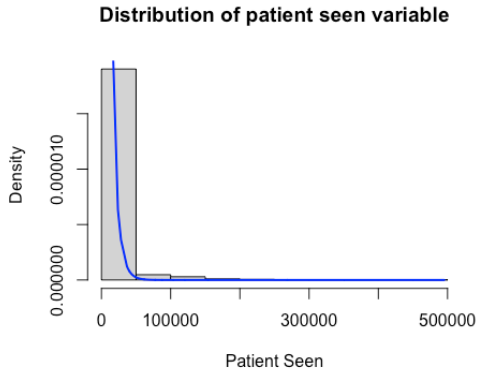
Data summary of numerical variables:

	count	mean	std	min	25%	50%	75%	max	unique
PATIENTS_SEEN	14535	7967.8	30243.2	4	1139	1774	2657	495141	2462
POPULATION	14535	15306.3	60077.2	0	2104	3000	4365	962934	2664
NUMBER_OF_DENTIST	14516	43.6	34.5	4	20	34	56.25	258	127
AVG_IMD_SCORE	14516	21.9	8.0	7.2	16.1	20.8	27.3	52.1	191

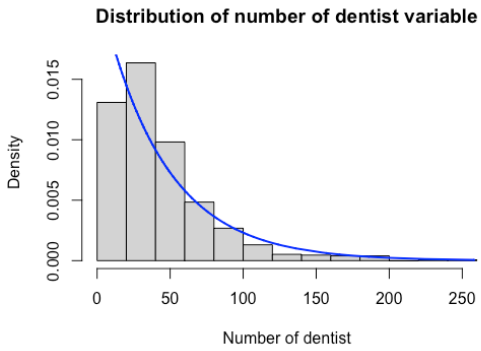
Below histogram charts shows distribution and normality of the data:

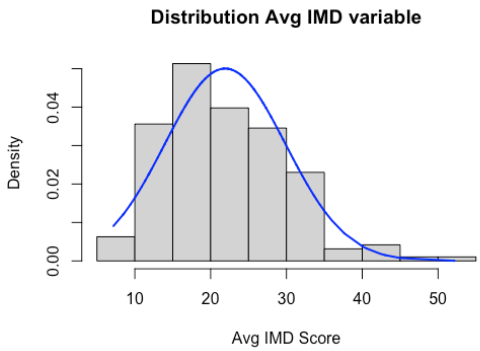
Population and “patient seen” variable is exponentially distributed and skewed to right tail



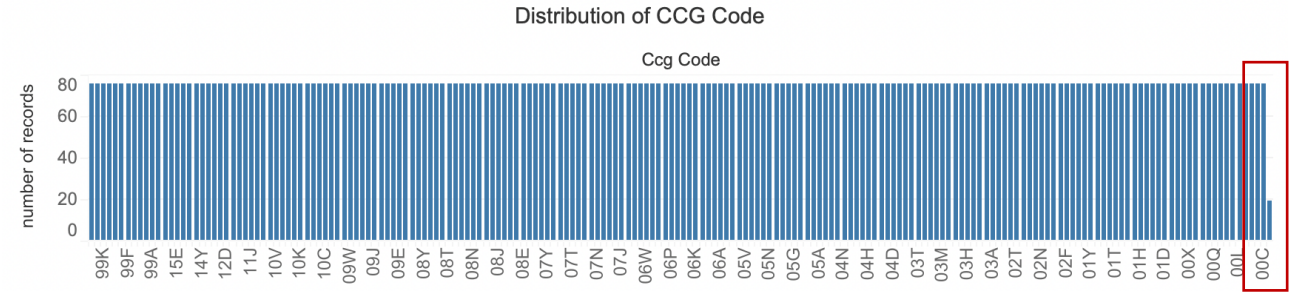


Dentist variable is exponentially distributed and Avg. IMD is normally distributed.





Data distribution for CCG code appears to be consistent across all CCG code (**80** records) except 1 code which is unallocated. Similar trend can be seen for CCG Name and CCG ONS Code.

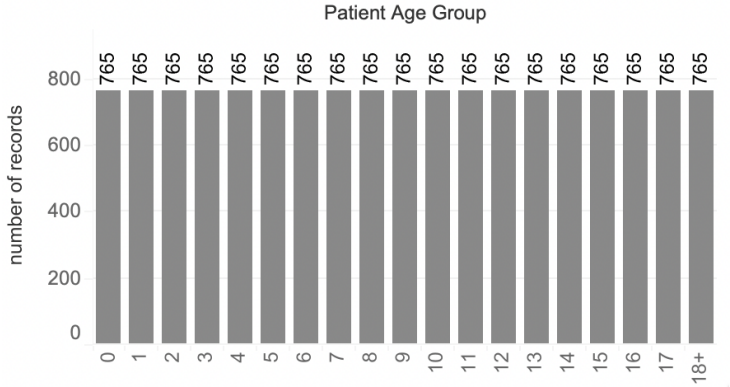


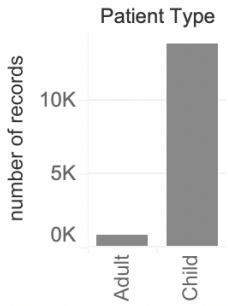
Record count for dentist age group variable seems to be consistent except few null values. Patient type variable has high number of records for children than adults which appears to be correct as children's data is more granular than adults in terms of grouping of age

Number of records for patient age group variable seems to be consistent.

Dentist age group - number of records

Dentist Age Group	
Null	19
35-44	3,629
45-54	3,629
55+	3,629
Under 35	3,629





c. Correlation

Correlation graph shows strong relationship between population and patient seen. However, there is no correlation between any variable with Avg. Deprivation score.
Additionally, number of dentists is weakly related with any other variables.

3. Data quality

1. Completeness :

Following three raw files are in scope for data quality checks:

File 1 Patient Seen: nhs-dent-stat-eng-jan-jun-19-anx3-ps-prac.csv
File 2 Patient Seen: nhs-dent-stat-eng-jul-dec-18-anx3-ps-prac.csv
Workforce File (Dentist data):NHS Dental Statistics for England 2018-19 Annex3_Workforce.csv
IMD File: File_13_-_IoD2019_Clinical_Commissioning_Group__CCG__Summaries

For patient seen dataset, there are 540 records that failed the data quality checks for completeness for the variable population. Other variables that fail this check are not considered for further analysis as they are not in scope for this report.

For workforce and IMD datasets, the variables in scope pass this quality check.

Patient Seen	Raw File 1	Raw File 2	Merged
PSEEN_END_DATE	0	0	0
GEOG_TYPE	0	0	0
PRACTICE_CODE	39102	39387	44004
PRACTICE_NAME	39102	39387	44004
PRAC_POSTCODE	39102	39387	44004
CCG_CODE	17214	17271	0
CCG_ONS_CODE	17214	17271	0
CCG_NAME	17214	17271	0
LA_CODE	21888	22116	44004
LA_NAME	21888	22116	44004
REGION_LO_CODE	17214	17271	0
REGION_LO_ONS_CODE	17214	17271	0
REGION_LO_NAME	17214	17271	0
REGION_CODE	17214	17271	0
REGION_ONS_CODE	17214	24168	6897
REGION_NAME	17214	24168	6897
PATIENT_TYPE	0	0	0
AGE_BAND	0	0	0
PATIENTS_SEEN	0	0	0
POPULATION	805272	808845	570

Workforce Files	Workforce File
Year	0
Geography	0
Parent_Code_1	0
Parent_Code_2	660
Org_Code	1089
Contract_Type	0
Age_Group	0
Gender	0
Dentist_Count	0

Integrated Dataset	Missing Values
PSEEN_END_DATE	0
CCG_CODE	0
CCG_ONS_CODE	0
CCG_NAME	0
PATIENT_TYPE	0
AGE_BAND	0
PATIENTS_SEEN	0
POPULATION	0
DENTIST_AGE_GROUP	19
NUMBER_OF_DENTIST	19
AVG_IMD_SCORE	19

Other data quality issues that are observed in the integrated dataset are as below:

- CCG code has value “unallocated” and shows patient seen count of 535348 without any population.

Ccg Code	Ccg Name	Ccg Ons Code	Patients Seen	Population
Unallocated	Unallocated	Unallocated	535,348	

- Dentist Age group contains **19 missing records** in final dataset as there is no age group for CCG with unallocated code.
- There are **570** missing records for population for CCG from patient seen merged files.
- Data is getting duplicated due to **cross join between patient age and dentist age** which is causing duplication of patient seen, population and dentist values.

2. **Correctness**: Dataset seems to pass this data quality check based on the trend and outlier analysis. It also verified with the source data on NHS website.

   - There are few outliers in the data such as population count of patient age group which can be justified as age group is segregated by patient type.
   - Patient seen data is correct however it is not in sync with the data period mentioned in the NHS data source as activity data available from 12 month to 31/03/2019 but patient seen data is from 30/Jun/2018 to 30/Sep/2019.
   - Population data has minimum value equals to 0 for 'Unallocated' CCG which appears to be data quality issue.

3. **Relevancy:**

   Data contains all the required attributes to complete detailed analysis of all the three question of the target report.

4. Detailed Analysis

This section covers the in-depth analysis of 12 months to 31-Mar-2019 NHS Dental data at CCG level to uncover findings that support the conclusion for below questions:

1. What patterns are there in the number/age of patients treated?

This part of analysis is to explore the relationship between number of patients and their age. The count for "Patients seen" variable for adult is cumulative to previous 24 months whereas for children it is cumulative to previous 12 months.

Since patient seen data is not on the same scale for adults and children, it is recommended to perform separate analysis for each patient type.

Analysis 1: Focus on numbers of child patient in different age group

Fig 4.2: Percentage of child patient vs child population

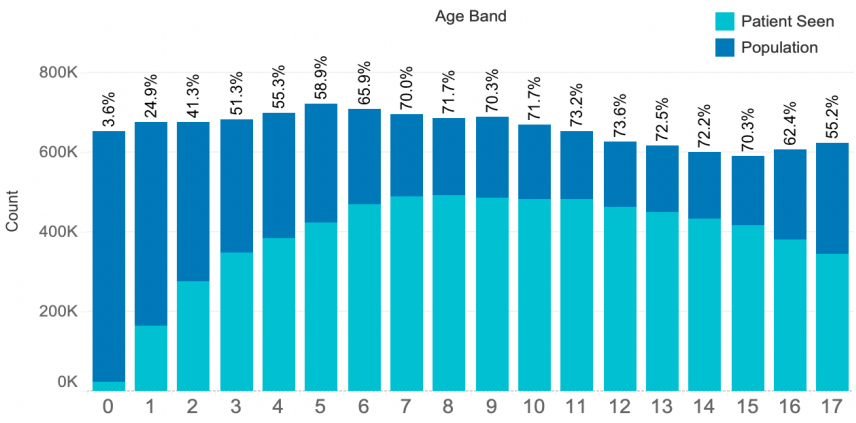


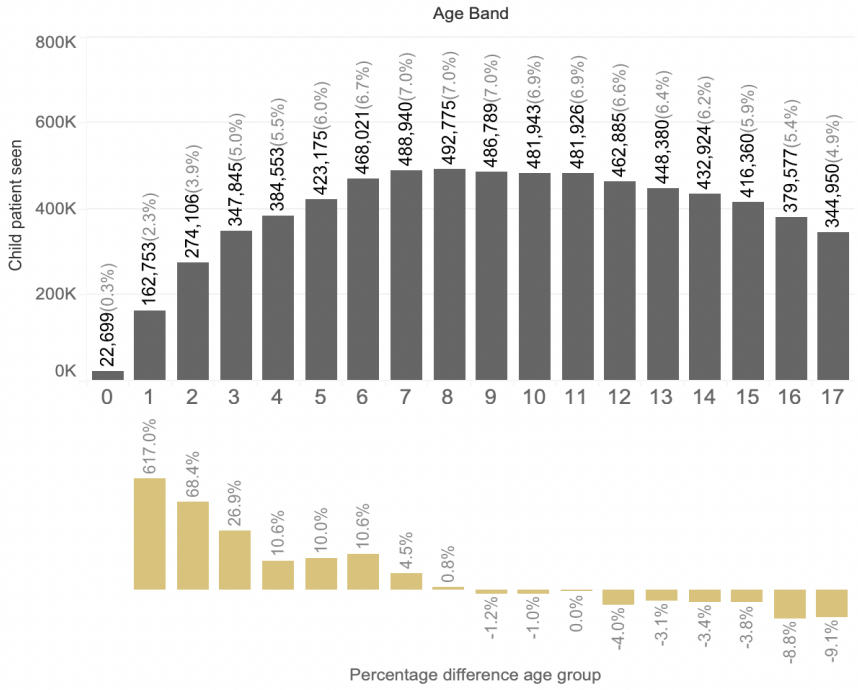
Figure 4.2 depicts insights about percentage of child patient in individual age group in comparison to overall population of the child in that group.

More than **50%** of children start visiting dentist by **age of 4** year which subsequently increases until 15 year age.

Figure 4.3 provides insights about how child patient volume is distributed across age groups and it is important to understand the child patient volume rate changes as per age group.

Fig 4.3: Percentage distribution of the child patient seen by age group

Data is analysed for patient seen end date Apr'19



From figure 4.3, it is evident that **~50% of child patients** are between the age group of **6 to 12 years**.

Post age of 9 years, number of child patients start declining.

In comparison to other age groups, child patient seen between 0 -1 years and 1 - 2 years have high difference (617% and 68.4%) which is expected as infant usually start getting teeth post 1 year which may trigger dentist visit.

Analysis 2: Number of adult patient across different age groups

As data suggest, there is single age group for adult which is 18+ hence only analysis supported with adult patient is comparative analysis with child population
Figure 4.4 provides insights that **76%** patient is in 18+ age group of overall patients numbers.
Figure 4.5 provides insights that **50.3%** of the total adult population visited dentist in last 24 month from April'19.

Fig 4.4 : Adult vs Child patient seen

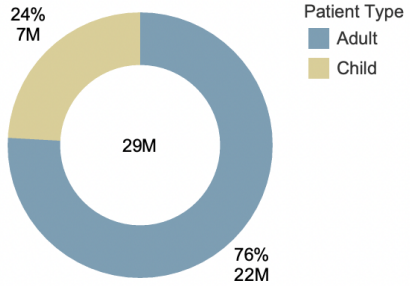
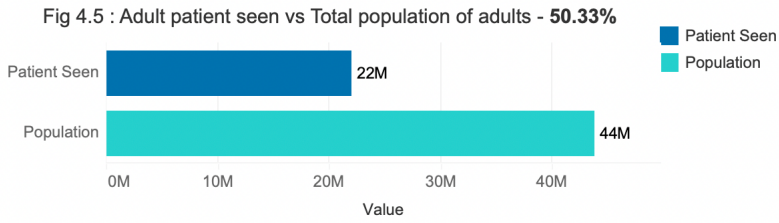


Fig 4.5 : Adult patient seen vs Total population of adults - **50.33%**



2. What is the effect of deprivation?

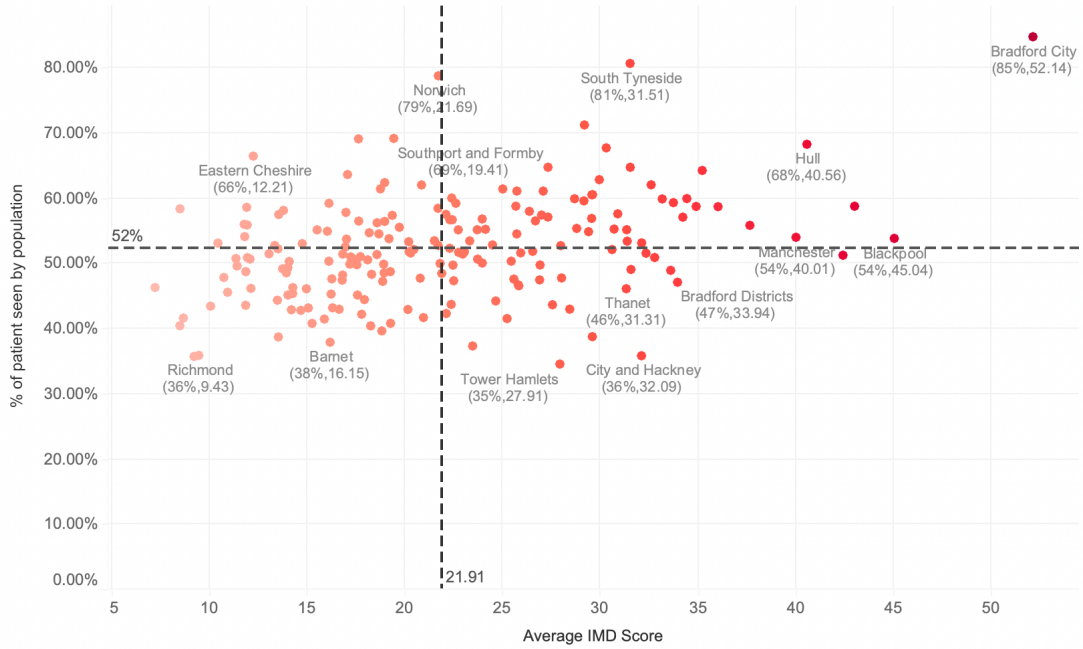
The Index of Multiple Deprivation (IMD) is the official measure of relative deprivation in England. Multiple components of deprivation are weighted with different strengths and compiled into a single score of deprivation. People may be considered to be living in poverty if they lack the financial resources to meet their needs, whereas people can be regarded as deprived if they lack any kind of resources, not just income.

This part of report is to analyse the effect of deprivation on the patients and dentist attributes across CCG and highlight any findings.

Below is the scatter plot between deprivation score and percentage of patient seen in individual CCG which provides insight about relationship between these two attributes.

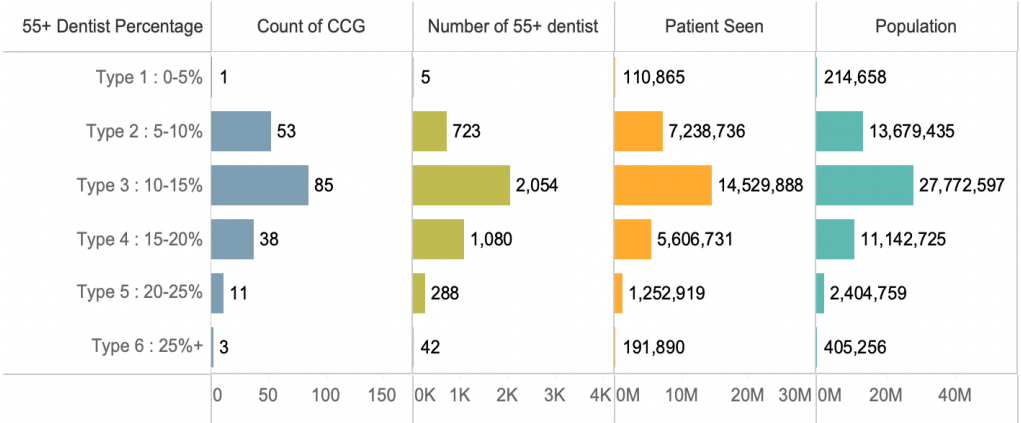
1. NHS Bradford City CCG shows highest deprivation score (52.14) with 85% of the population visited the dental care.
2. It appears that if you have lower IMD than average IMD score (22) of all CCG, then it is more likely that % of patient seen is low (below 52% average % of patient seen by population)
3. There are few highlights such as NHS Norwich shows higher number of patient even it IMD score is considerably lower in comparison to others.
4. There are few CCG such as NHS City and Hackney CCG, Liverpool who has lower % of patient even though they have IMD score, need further analysis as other deprivation domains might causing higher IMD score.

Fig 4.6 : Scatter plot to between Avg. IMD score and percentage of patient seen by total poplation.
Each dot represent the CCG on the scale of % of patient by population and Avg. IMD Score



3. A person aged 55 in 2019 will reach the State Pension age in 2031. What types of CCGs face the greatest shortage of dentists in 2031?
We can define **the type of CCG by bucketing the percentage of the 55+** dentist in individual CCG.

Fig 4.7: Distribution of attributes by percentage of 55+ age dentist in CCG
(55+ Dentist percentage is the percentage of dentist above or equal to 55 Age on April' 2019 in CCG)



Bar chart plotted on 2019 data, hence dentist in 55+ age group would retired by 2031.

3 CCGs in Type- 6 25%+ are most impacted as 25% of dentist in workforce will retire by 2031.

Type 4 & 5 CCG's (25% of overall CCG's count) **will lose their 15-25%** of workforce by 2031 which is also a significant number impacting population of 13M. **Type 3 CCG's will lose 10-15%** which does not look significant impact however in terms of population and patient seen it might be impacting as it is ~50% of overall population and patient seen data.

Scatter plots help us to understand the spread of the risked CCG's. Values shown in red are more risked as they have significant population of 55+ dentist retiring by 2031 whereas the values in amber are less and values in green are least impacted.

Fig 4.8 : Scatter plot to show CCG spread across population and % of 55+ aged dentist with RGB
(Assumption: Red: High impact , Amber: Medium Impact, Green: Low Impact)

5.Conclusion:

As per data characterisation and data quality check results, NHS data showed more than 98% completeness and correctness which increased the relevancy of the analysis. From this report we are able to support the analysis performed on age group of patients, dentist and deprivation scores along with the key insights using visualizations.

As examined from patient seen and age group analysis, child patient showed an upward trend across the age group, more than 50% of child population started visiting dentist by age of 4 years and subsequently kept on increasing until the age group of 9 years post which it started declining.

As a part of analysis, interesting facts are uncovered after joining deprivation and patient seen data, such as higher deprivation score might lead to higher percentage of patients which is supported by significant number of CCG's as shown in scatter plot (Fig 4.6). Though there are few CCG's which do not sync with this analysis and need further rigorous analysis to find exact reason for deprivation.

Additionally, NHS need to be cautious to maintain the ratio of dentist across age groups. Imbalance in dentist count across age group might affect the dentist count versus patient ratio in future. As per the analysis in this report, more than 30% CCG's will have dentist crisis as their 20% workforce retire by 2031.