Gian Clarke

GRC0396

Task

Cleansing data to create a functional database

Assignment 1

Hospital Patient Waitlist

**Data Cleansing**

**Doctors**

Table

Description automatically generated

* Replaced all duplicate values and assigned a unique DoctorNo to each doctor
* Used: Data > Remove Duplicates

**Referral**

Graphical user interface, application

Description automatically generated

* Replaced the Referred From doctor name with the new DoctorNo
* Used: CTR + H > Replace All
* HealthTargetEligible, changed yes and no to integer 1 = yes, 0 = no. this saves space and less time to query

**Patient**

Table

Description automatically generated





* 6 Patients have the same NHI which is meant to be a unique value
* Used: Replaced three of the exact matches to create new unique IDS
* Removed Wandis Clipson and Graeme Fenemore because they are time travellers
* Split first and last name into separate fields to make it easier to query a person by name

**Department**

Graphical user interface, text, application, chat or text message

Description automatically generated

* Replaced duplication departments and assigned an ID to each department
* Used: Data > Replace Duplicates

**Surgeons**

Table

Description automatically generated

* Removed duplicate surgeons and added ID for each and referenced DepartmentNo against each doctor
* Used: Data > Replaced Duplicates, CTR + H Replace All

**The Design Process**

1. **Split the data provided into 5 database tables**

* During the process of cleansing the data, I found there to be 3 patients with incorrect dates and one being that they were born prior to their referral date, these three clients got deleted as the data was wrong and could not contact them to get the correct date of birth. The other two had ridiculous date of births.
* Six clients had duplicate NHI codes which I believe they should be unique, and this was human error, as all the clients were unique apart from their NHI numbers I decided to keep them as they are real patients and not duplicate ones. I changed the duplicate NHI of each client and made sure no more duplicates existed in the database.

1. **Created a ERD Diagram to show the relationship of tables**

* Initially created a ERD diagram with six tables, after starting to create the tables in SQL I quickly realised that there was no point in having a referral table and a waitlist table. I combined both tables into referral as there was so much duplicate information in waitlist and more relationships than required.

1. **Created required queries to display correct data**

* Many of the first queries created did not work for all required output, once I did further research, I found many other functions to create calculated data fields that you don’t have to manually add and update to a database.