CMP406 Task 1

2003115 Robbie Sangster

# Database Preparation

1. For data managing and processing Hive is used and imported into the CLI

A close-up of a computer screen

Description automatically generated

1. Create a table and load in the data from the csv file.

A computer screen shot of a computer code

Description automatically generated

1. Create a unique ID for each row by creating a new table and using the row number as the id.

A close-up of a computer screen

Description automatically generated

1. Run a select statement to check that data has inserted and is the correct format. The final column is the primary key.

A close-up of a document

Description automatically generated

# Database Analysis

This is where we can begin to analyse the staff salaries.

1. Staff who are paid more than their grade is worth.

SELECT \* FROM (SELECT id,Role,Annual\_Salary, AVG(Annual\_Salary) OVER (PARTITION BY Agency) AS Avg\_Salary FROM ACMEv2) t WHERE Annual\_Salary > Avg\_Salary ;

A number on a white background

Description automatically generated

(Full list in sperate file as 1454 rows fetched.)

1. Number of staff working full and part time.

SELECT FULL\_PART\_TIME AS value, COUNT(FULL\_PART\_TIME) FROM ACMEv2 GROUP BY Full\_Part\_Time;



1. Number of staff in each service type.

SELECT Service AS Service, COUNT(Service) FROM ACMEv2 GROUP BY SERVICE;

A white background with black text

Description automatically generated

1. Average salary in each service type.

SELECT Service AS Service, AVG(Annual\_Salary) FROM ACMEv2 GROUP BY SERVICE;

A close-up of numbers

Description automatically generated

1. Number of staff in each job role.

SELECT Role AS Role, COUNT(Role) FROM ACMEv2 GROUP BY Role;

A screenshot of a computer program

Description automatically generated

(Full output in separate file as 1401 rows collected.)

1. Staff earning below average salary for the university.

SELECT \* FROM (SELECT id,Role,Annual\_Salary, AVG(Annual\_Salary) OVER (PARTITION BY Agency) AS Avg\_Salary FROM ACMEv2) t WHERE Annual\_Salary < Avg\_Salary;

A number with numbers on it

Description automatically generated with medium confidence