SQL-Mongo Project – Home Credit Data

BUAN 6320

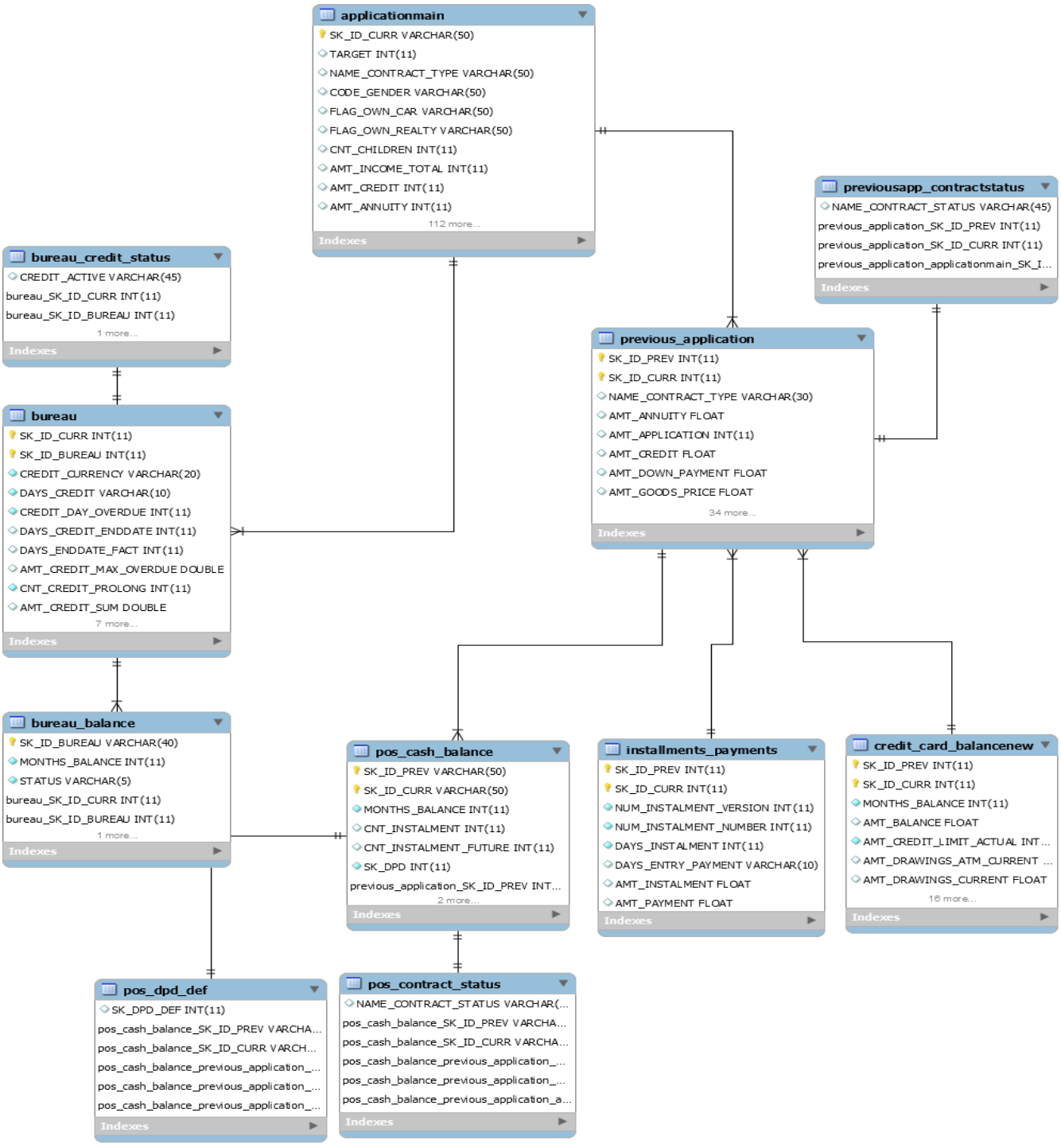
Evan Jones Boddu

Contents

Data Model

Assumptions/Notes About Data Entities and Relationships

Entity-Relationship Diagram



Physical Database

Assumptions/Notes About Data Set

Screen shot of Physical Database objects

Data in the Database

Table Name	Primary Key	Foreign Key	# of Rows in Table
applicationmain	SK_ID_CURR	-	356255
bureau	SK_ID_BUREAU	SK_ID_CURR	1716428
bureau_credit_status	SK_ID_BUREAU	SK_ID_CURR	1048575
bureau_balance	SK_ID_BUREAU	SK_ID_CURR	27299925
previous_application	SK_ID_PREV	SK_ID_CURR	1670214
previousapp_contractstatus	SK_ID_PREV	SK_ID_CURR	1048575
POS_Cash_balance	SK_ID_PREV	SK_ID_CURR	10001358
POS_contract_status	SK_ID_PREV	SK_ID_CURR	1048575
POS_dpd_def	SK_ID_PREV	SK_ID_CURR	1048575
Installments_payments	SK_ID_PREV	SK_ID_CURR	13605401
credit_card_balancenew	SK_ID_PREV	SK_ID_CURR	3840312

SQL Queries

Query 1

Question:

Which occupation types had the highest and lowest number of loans approved?

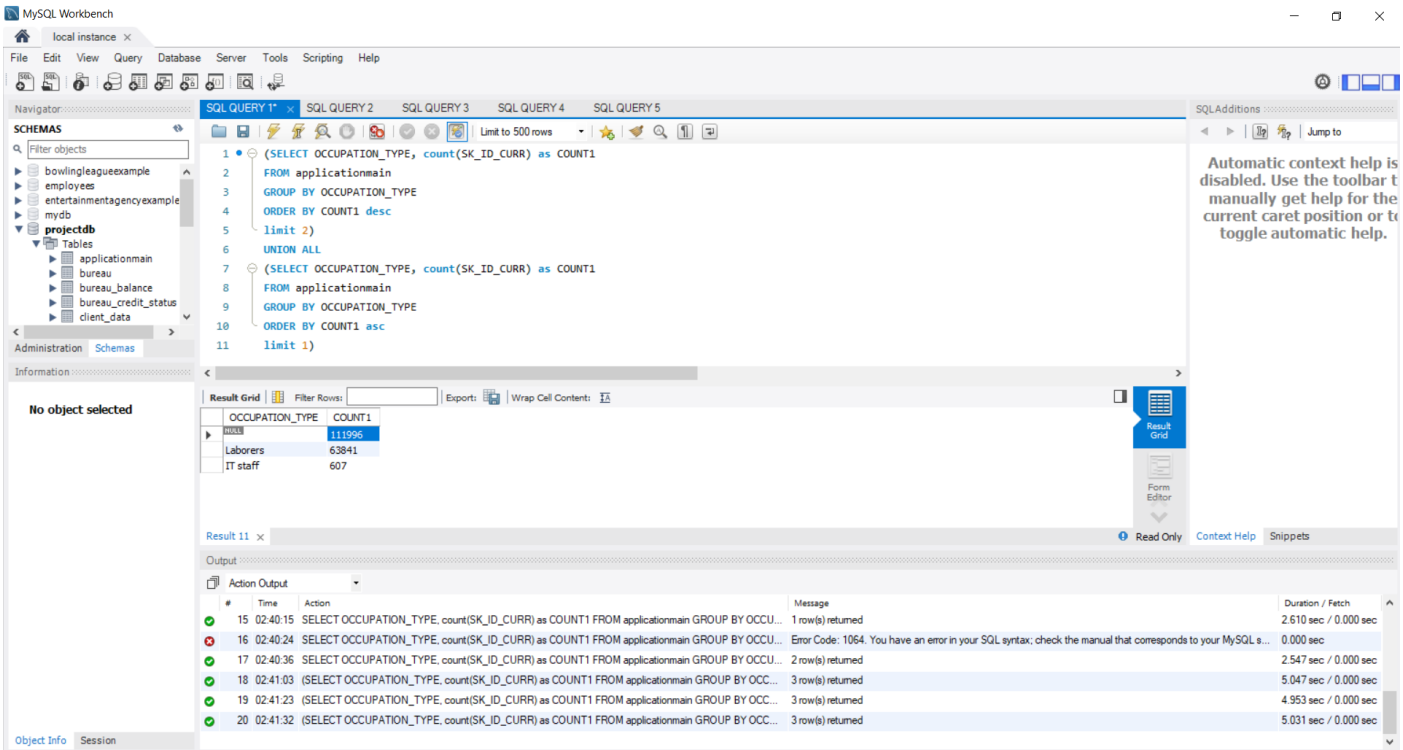
Notes/Comments About SQL Query and Results (Include # of Rows in Result)

From the occupation type, the highest number of loans approved are for Laborers which are 63841 and the least number of loans approved are for IT Staff which are 607.

Translation

SELECT OCCUPATION_TYPE, and GROUP BY the desc order (highest) and asc order (lowest) number of loans approved.

Screen Shot of SQL Query and Results



COUNT of Laborers (the highest) is 63841

COUNT of IT Staff (the lowest) is 607.

Query 2

Question:

What was the average loan amount approved for Revolving Loans?

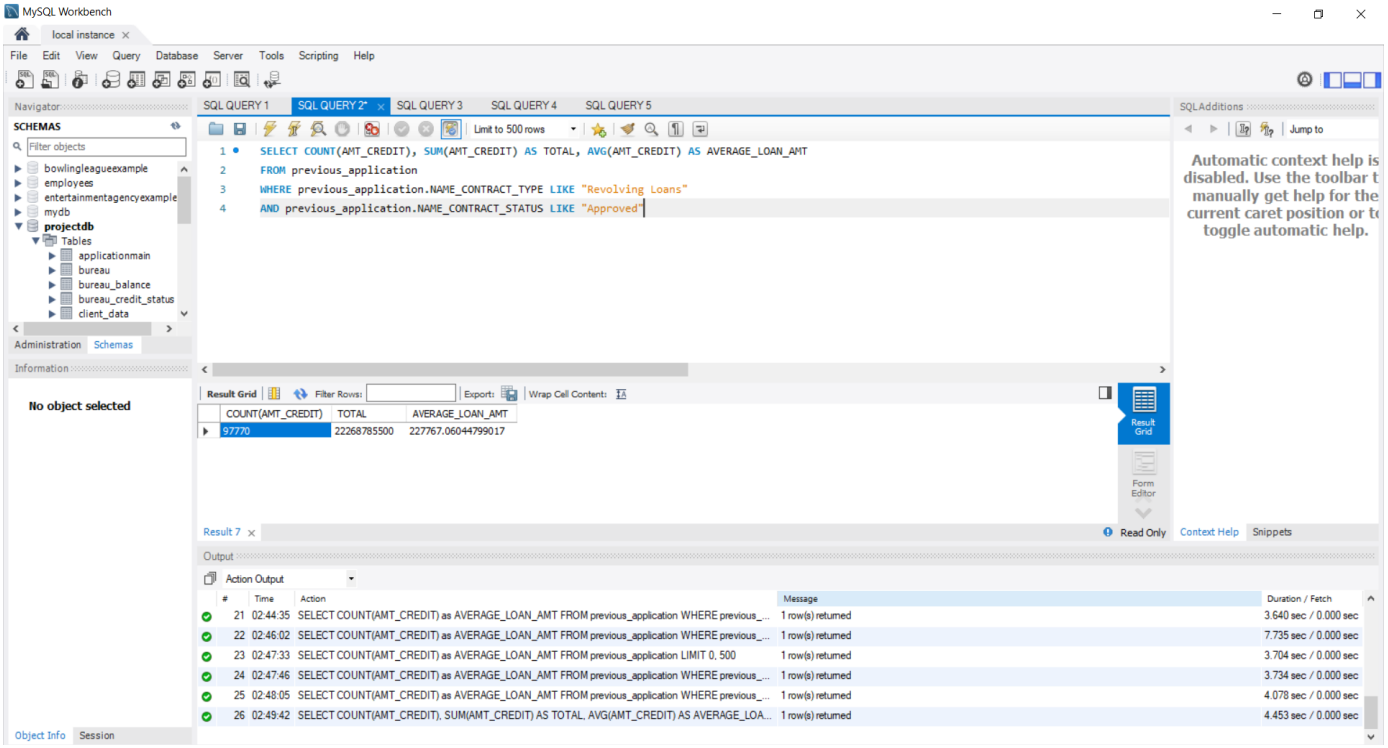
Notes/Comments About SQL Query and Results (Include # of Rows in Result)

The above code is done using previous_application. Average loan amount is 27767.0604.

Translation

SELECT AVG(AMT_CREDIT) who have been approved for Revolving Loans.

Screen Shot of SQL Query and Results



The Average Loan Amount Approved for Revolving loans is 27767.0604. Where the Total is 22268785500 and the count (i.e the number of observations for Revolving loans which are approved) are 97770.

Query 3

Question:

Did applicants with Higher education apply for more Revolving Loans than Cash Loans?

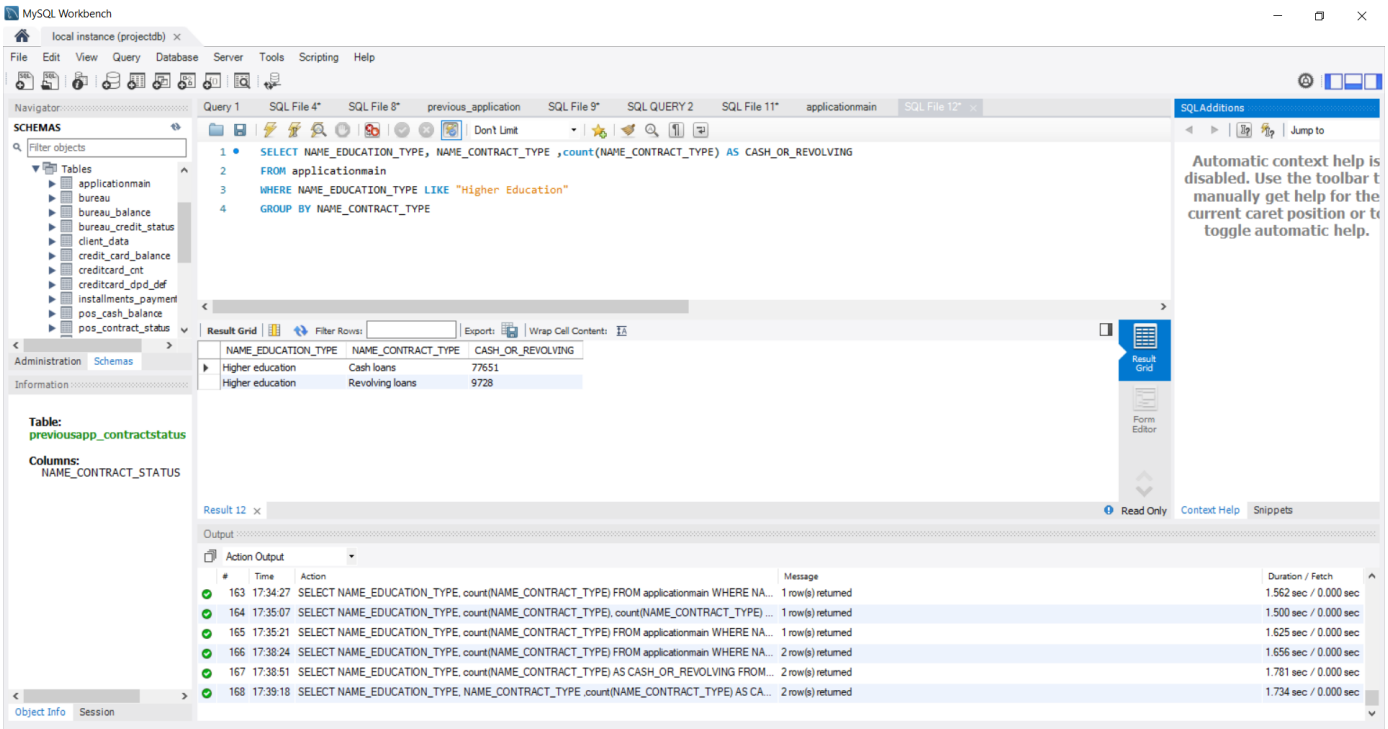
Notes/Comments About SQL Query and Results (Include # of Rows in Result)

No, the applicants who has higher education applied more for cash loans rather than revolving loans. As, the number of applicants who have higher education and applied for cash loans are 77651 and the number of applicants who have higher education and applied for revolving loans are 9728.

Translation

SELECT NAME_EDUCATION_TYPE, NAME_CONTRACT_TYPE, COUNT(NAME_CONTRACT_TYPE) who have higher education and the count of applicants who applied for cash loans and revolving loans.

Screen Shot of SQL Query and Results



Applications having higher education and cash loans are more than applications having higher education and revolving loans. Cash Loan applicants are 77651, while the Revolving loan applicants are 9728.

Query 4

Question:

What was the average credit utilization percentage for active (not repaid) loan applicants?

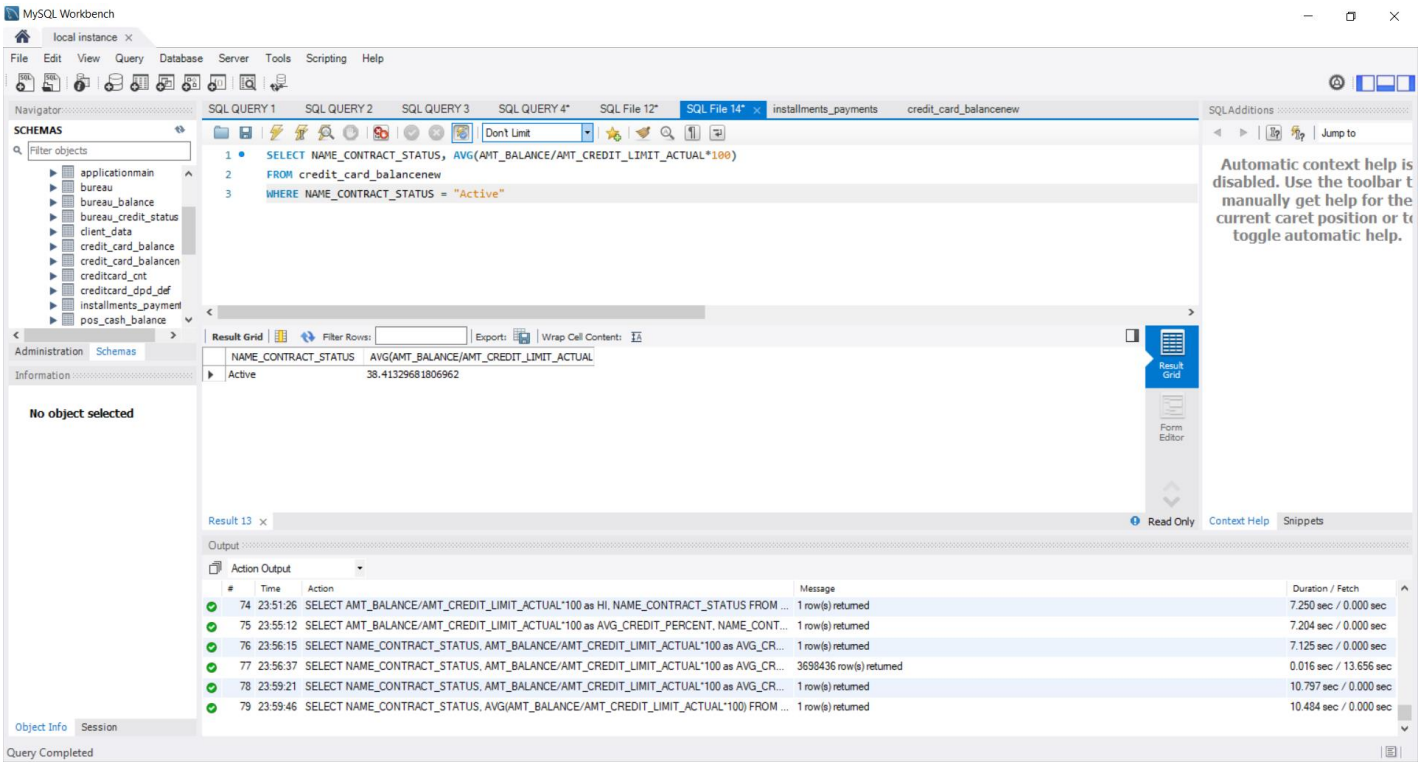
Notes/Comments About SQL Query and Results (Include # of Rows in Result)

The average credit utilization percentage for active(unpaid) loan applicants is 38.4133.

Translation

SELECT NAME_CONTRACT_STATUS, AVG(AMT_BALANCE/AMT_CREDIT_LIMIT_ACTAL*100) where the NAME_CONTRACT_STATUS is active.

Screen Shot of SQL Query and Results



The Average credit utilization percentage for active loan applicants is 38.4133%

Query 5

Question:

Applicants of which education type had the highest average unused credit?

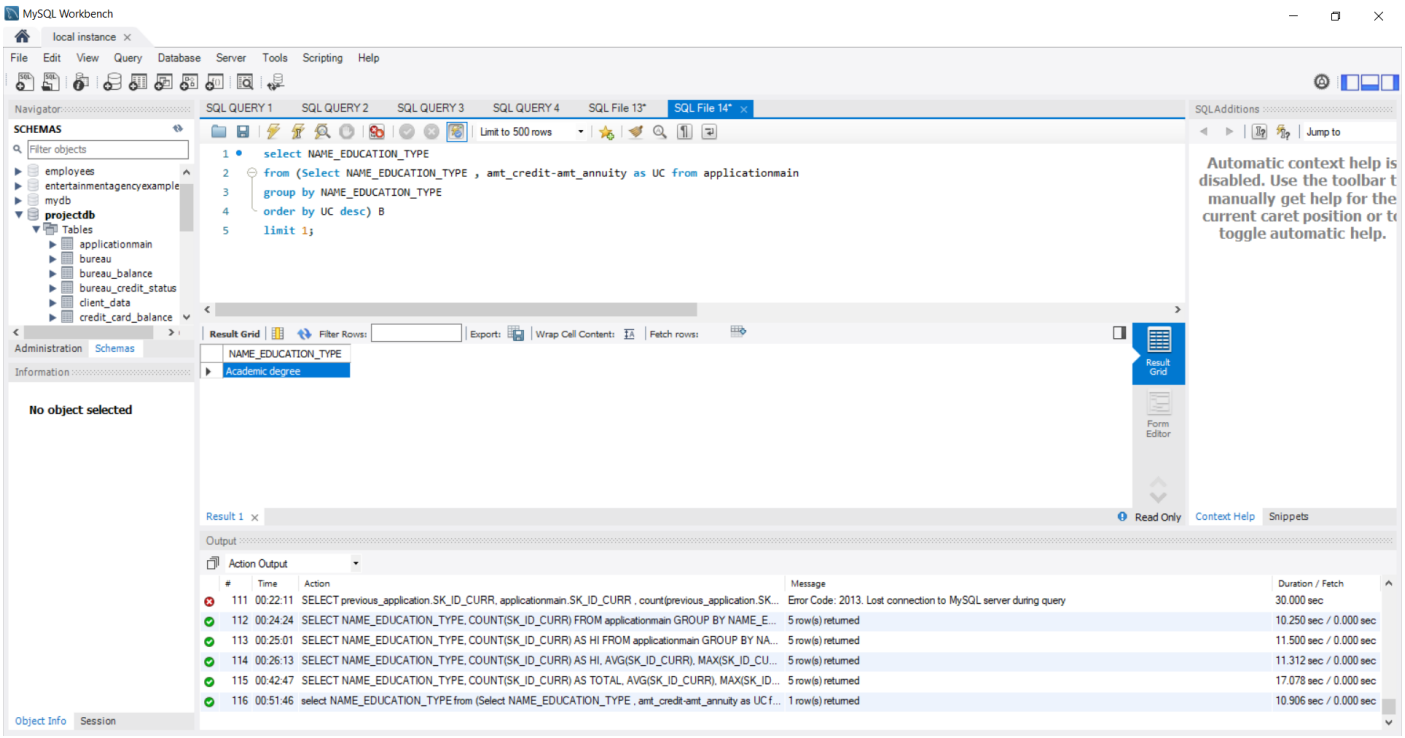
Notes/Comments About SQL Query and Results (Include # of Rows in Result)

The Applicant of which education type had the highest average unused credit is Academic Degree.

Translation

SELECT NAME_EDUCATION_TYPE where the applicant had the highest average unused credit.

Screen Shot of SQL Query and Results



The Applicant of which education type had the highest average unused credit is Academic Degree.