**School of Business | University of Kansas**

**BSAN440/IST495 Foundations in Business Analytics**

**Due to large print out for queries, screenshot of output was limited to 10 rows or less.**

1. (20 pts) Complete the exercise problem from the lab notes.
   a.) Find the average imdb score for movies release between 2000 and 2010.

| | ROUND(AVG(IMDB_SCORE),2) |
|---|---|
| 1 | 6.35 |

   b.) Get the title year and count of films released in each year.

| | TITLE_YEAR | COUNT(TITLE_Y... |
|---|---|---|
| 1 | 1916 | 1 |
| 2 | 1920 | 1 |
| 3 | 1925 | 1 |
| 4 | 1927 | 1 |
| 5 | 1929 | 2 |
| 6 | 1930 | 1 |
| 7 | 1932 | 1 |
| 8 | 1933 | 2 |
| 9 | 1934 | 1 |
| 10 | 1935 | 1 |

   c.) Get the title year and the largest budget for all movies, grouped by title year.

| | TITLE_Y... | MAX(BUDGET) |
|---|---|---|
| 1 | 2009 | 250000000 |
| 2 | 2011 | 250000000 |
| 3 | 2014 | 250000000 |
| 4 | 1999 | 1000000000 |
| 5 | 1990 | 100000000 |
| 6 | 1962 | 15000000 |
| 7 | 1933 | 439000 |
| 8 | 1957 | 3000000 |
| 9 | 1936 | 1500000 |
| 10 | 1939 | 3977000 |

   d.) Get the title year and lowest gross earnings per title year.

| | TITLE_Y... | MIN(GROSS) |
|---|---|---|
| 1 | (null) | 145118 |
| 2 | 2016 | 31662 |
| 3 | 2015 | 1711 |
| 4 | 2014 | 162 |
| 5 | 2013 | 2843 |
| 6 | 2012 | 1332 |
| 7 | 2011 | 2245 |
| 8 | 2010 | 12561 |
| 9 | 2009 | 828 |
| 10 | 2008 | 3478 |

   e.) Get the language and total gross amount films in each language made.

| | LANGUAGE | COUNT(GROSS) | TOTAL GROSS AMOUNT |
|---|---|---|---|
| 1 | Hindi | 21 | 46559734 |
| 2 | Kazakh | 1 | 77231 |
| 3 | Cantonese | 10 | 64294253 |
| 4 | German | 15 | 43748635 |
| 5 | Swahili | 0 | (null) |
| 6 | Mandarin | 18 | 163611530 |
| 7 | Maya | 1 | 50859889 |
| 8 | Polish | 3 | 4720641 |
| 9 | Thai | 3 | 12461829 |
| 10 | Danish | 3 | 2403857 |

   f.) Get the title year, country, and highest budget spent making a film for each year, for each country. Sort your results by release year and country.

| | TITLE_YEAR | COUNTRY | MAX(BUDGET) |
|---|---|---|---|
| 1 | 2003 | Afghanistan | 46000 |
| 2 | 2009 | Argentina | 2000000 |
| 3 | 2004 | Argentina | 1400000 |
| 4 | 2000 | Argentina | 1500000 |
| 5 | 1998 | Aruba | 35000000 |
| 6 | 2015 | Australia | 150000000 |
| 7 | 2014 | Australia | 65000000 |
| 8 | 2013 | Australia | 105000000 |
| 9 | 2012 | Australia | 30000000 |
| 10 | 2011 | Australia | 135000000 |

g.) Get the title year, country, and highest gross of a film for title year after 2000.

| | TITLE_YEAR | COUNTRY | MAX(GROSS) |
|---|---|---|---|
| 1 | 2009 | USA | 760505847 |
| 2 | 2016 | USA | 407197282 |
| 3 | 2012 | UK | 304360277 |
| 4 | 2016 | UK | 362645141 |
| 5 | 2001 | New Line | 6712451 |
| 6 | 2001 | New Zealand | 313837577 |
| 7 | 2015 | China | 195000874 |
| 8 | 2011 | Australia | 63992328 |
| 9 | 2005 | Germany | 82931301 |
| 10 | 2009 | Spain | 42194060 |

h.) Get the title year, average budget and average gross earnings for title year after 2000. Alias the average budget as avg_budget; alias the average gross earnings as avg_gross.

| | TITLE_Y... | AVG_BUDGET | AVG_GROSS |
|---|---|---|---|
| 1 | 2001 | 37687307.69 | 44595685.11 |
| 2 | 2002 | 32598510.69 | 44841235.69 |
| 3 | 2003 | 37208648.83 | 51562598.3 |
| 4 | 2004 | 46865344.62 | 43255326.25 |
| 5 | 2005 | 70323938.23 | 43128699.22 |
| 6 | 2006 | 93549448.6 | 41816170.87 |
| 7 | 2007 | 35117483.7 | 49573380.02 |
| 8 | 2008 | 41804885.57 | 47940413.81 |
| 9 | 2009 | 37073287.04 | 51711597.7 |
| 10 | 2010 | 45476979.7 | 55341567.53 |

i.) Modify your query so that only years with an average budget of greater than $20 million are included.

| | TITLE_YEAR | AVG_BUDGET | AVG_GROSS |
|---|---|---|---|
| 1 | 2001 | 55800446.43 | 63403471.13 |
| 2 | 2002 | 56029702.97 | 70750112.54 |
| 3 | 2003 | 62496385.54 | 82243840.43 |
| 4 | 2004 | 76046249.09 | 59353095.79 |
| 5 | 2005 | 121366071.43 | 62058593.9 |
| 6 | 2006 | 193826020.41 | 69879008.84 |
| 7 | 2007 | 69443750 | 83590732.28 |
| 8 | 2008 | 68070290.91 | 71083687.98 |
| 9 | 2009 | 68387128.71 | 88050075.48 |
| 10 | 2010 | 83092783.51 | 85609660.73 |

2. (20 pts) Import data file *CustomerBasic.cs*v to the database through SQL Developer.
   **a.** How many rows in this table?

| | NUMBER OF ROWS |
|---|---|
| 1 | 5000 |

   b. How many males and how many females in the table?

| | GENDER | COUN... |
|---|---|---|
| 1 | Male | 2473 |
| 2 | Female | 2494 |
| 3 | (null) | 33 |

   c. How many customers whose gender is missing? **SEE ABOVE OUTPUT**
   d. What is the average and standard deviation of age for male and female, respectively? Keep two decimals.

| | GENDER | ROUND(AVG(AGE),2) |
|---|---|---|
| 1 | Male | 47.2 |
| 2 | Female | 46.87 |
| 3 | (null) | 45.21 |

e. How many job categories, and what is their frequency (change the column name to FREQ)? Sort the resulted table by the frequency in descending order.

| | JOBCATEGORY | FREQ |
|---|---|---|
| 1 | (null) | 0 |
| 2 | Service | 620 |
| 3 | Sales | 1635 |
| 4 | Professional | 1380 |
| 5 | Labor | 686 |
| 6 | Crafts | 452 |
| 7 | Agriculture | 212 |

f. What is the frequency for married and unmarried customers in the age between 20 and 30 (included)?

| | MARITALSTATUS | FREQ |
|---|---|---|
| 1 | Unmarried | 499 |
| 2 | Married | 476 |

3. (20 pts) Import data file *CustomerFin.csv* to the database through SQL Developer.
   a. **Select columns CUSTOMERID, GENDER, AGE, JOBCATEGORY from** CustomerBasic table and HHINCOME, DEBTTOINCOMERATIO, LOANDEFAULT, CARDTENURE from CustomerFin table, and inner join together, and create a new table named CustomerAll.

| | COLUMN_NAME | DATA_TYPE | NULLABLE | DATA_DEFAULT | COLUMN_ID | COMMENTS |
|---|---|---|---|---|---|---|
| 1 | CUSTOMERID | VARCHAR2(26 BYTE) | Yes | (null) | 1 | (null) |
| 2 | GENDER | VARCHAR2(26 BYTE) | Yes | (null) | 2 | (null) |
| 3 | AGE | NUMBER(4,0) | Yes | (null) | 3 | (null) |
| 4 | JOBCATEGORY | VARCHAR2(26 BYTE) | Yes | (null) | 4 | (null) |
| 5 | HHINCOME | NUMBER(9,0) | Yes | (null) | 5 | (null) |
| 6 | DEBTTOINCOMERATIO | NUMBER(5,1) | Yes | (null) | 6 | (null) |
| 7 | LOANDEFAULT | VARCHAR2(26 BYTE) | Yes | (null) | 7 | (null) |
| 8 | CARDTENURE | NUMBER(4,0) | Yes | (null) | 8 | (null) |

   b. What is the average (keep two decimals), minimum and maximum income for each job category without missing jobs? Note that the result should include job category.

| | JOBCATEGORY | ROUND(STDDEV(HHINCOME),2) | MAXINCOME | MININCOME |
|---|---|---|---|---|
| 1 | Sales | 36488.33 | 780000 | 9000 |
| 2 | Labor | 75558.95 | 995000 | 9000 |
| 3 | Professional | 47015.09 | 642000 | 9000 |
| 4 | Service | 72643.9 | 1073000 | 9000 |
| 5 | Crafts | 62677.97 | 575000 | 9000 |
| 6 | Agriculture | 60942.34 | 472000 | 9000 |

   c. Obtain the frequency of defaulted customer by job category (without missing).

| | JOBCATEGORY | FREQ OF DEFAULTED CUSTOMERS |
|---|---|---|
| 1 | Agriculture | 25 |
| 2 | Crafts | 64 |
| 3 | Labor | 90 |
| 4 | Service | 101 |
| 5 | Agriculture | 146 |
| 6 | Professional | 287 |
| 7 | Crafts | 305 |
| 8 | Sales | 376 |
| 9 | Service | 384 |
| 10 | Labor | 453 |

   d. What is the default rate for each job category? (You can use excel or calculator to get the results if you don't know how to write the SQL.)

| | JOBCATEGORY | FREQ OF DEFAULTED CUSTOMERS |
|---|---|---|
| 1 | Sales | 935 |
| 2 | Labor | 453 |
| 3 | Professional | 826 |
| 4 | Professional | 287 |
| 5 | Crafts | 64 |
| 6 | Service | 384 |
| 7 | Labor | 90 |
| 8 | Crafts | 305 |
| 9 | Agriculture | 146 |
| 10 | Sales | 376 |