Faculty of Information Technology

BICS

Advanced Database

Semester Project

_____

Area: Data Warehousing, Extraction Transformation & Loading and Data Visualization	**Tools:** Talend Studio for Data Intergration, Postres/MySQL/Oracle Database, Tableau, Any scripting language (Preferably Python, JavaScript or PHP)

Narrative

Kiva is an international nonprofit, founded in 2005 and based in San Francisco, with a mission to connect people through lending to alleviate poverty. We celebrate and support people looking to create a better future for themselves, their families and their communities. By lending as little as $25 on Kiva, anyone can help a borrower start or grow a business, go to school, access clean energy or realize their potential. For some, it's a matter of survival, for others it's the fuel for a life-long.

The life cycle of a KIVA loan is as follows: A borrower applies for a loan, the loan goes through the underwriting and approval process, then the load undergoes Loan disbursal period (i.e. when the borrower can access the money— the timing of this can vary), after which the borrower repays the loan and finally Lenders use repayments to fund new loans, donate or withdraw the money.

Project Instructions

You have been approached by the board of the KIVA loan to come up with a modern data warehouse. The board seeks to finally have detailed analytics on the following areas:

  i.   The borrowing rate across different times of the year
 ii.   The borrowing trend across different regions
iii.   The type of loans mostly applied for
 iv.   The disbursement rate across the year
  v.   The fundraising trend across different regions in comparison to the time of the year

vi.   The repayment trend in the different regions with the passing of time
vii. Reports of expenses incurred in the course of different KIVA loan cycles
viii. The demographic trend of the borrower across different regions in relation to time of the year
ix.   The demographic trend of the fundraisers across different regions in relation to time of the year

All that has been provided for you are datasets which are attached together with this document. You are expected to study the datasets and come up with different data marts for each group. If need be, where the dataset is not sufficient, you are expected come up with additional arbitrary but meaningful datasets. Be keen to ensure you do not repeat a dimension or fact from other groups. Remember a data mart consists of dimensions and facts that presuppose a common business logic.

Upon completion, merge the data marts to form a data warehouse and use the data within this data warehouse to provide the KIVA board with the above reports by use of Tableau as your ETL tool and publish your workbooks if need be to Tableau online platform.

The allocation of reports per group are as follows:

Groups	Reports to generate(From the above list)
1 to 5	i, iv & ix
6 to 10	ii ,v &ix
11 to 15	iii ,vi &vii
16 to 22	viii , vii &ii

Project Deliverables

		Date:
1	Identification of arbitrary dimensions and facts & Design of star schemas,	Group A: 13/8/2021 Group B:17/8/2021
2	Populating arbitrary csv files with data and creation of data marts in Postgres/MySQL/Oracle	Group A:16/08/2021 Group B:17/08/2021
3	ETL	Group A:16/08/2021 Group B:17/08/2021
4	Linking of data marts and Visualization	Group A:20/08/2021 Group B:19/08/2021
Final Presentation	Mashup of all the above	Group A:23/08/2021 Group B:24/08/2021

Consultation Hours: Fridays 8:00am – 5:00pm

Email Address: jkirui@strathmore.edu