data = LOAD '/user/maria\_dev/train\_Loan.csv' USING PigStorage (',') AS (Loan\_ID:chararray, Gender:chararray, Married:chararray, Dependents:chararray, Education:chararray, Self\_Employed:chararray, ApplicantIncome:int, CoapplicantIncome:int, LoanAmount:int, Loan\_Amount\_Term:int, Credit\_History:int, Property\_Area:chararray, Loan\_Status:chararray);

--Number of Male who got loans approved

male\_loan = Filter data by Gender == 'Male' and Loan\_Status == 'Y';

male\_loan\_group = Group male\_loan All;

count = Foreach male\_loan\_group Generate COUNT(male\_loan);

dump count;

-- Number of females who got loan approved

female\_loan = Filter data by Gender == 'Female' and Loan\_Status == 'Y';

female\_loan\_group = Group female\_loan All;

count = Foreach female\_loan\_group Generate COUNT(female\_loan);

dump count;

--Maximum Loan amount term for approved loans

approved\_loan = Filter data by Loan\_Status == 'Y';

approved\_group = Group approved\_loan All;

max = foreach approved\_group generate MAX(approved\_loan.Loan\_Amount\_Term);

dump max;

--Top property area for highest approved loan amount;

loan\_order = Order data by LoanAmount DESC;

highest\_loan = Limit loan\_order 1;

dump highest\_loan;

--Which group has a higher loan approval rate from the below ?

--Male vs Female ?

approved\_loan = Filter data by Loan\_Status == 'Y';

gender\_group = Group approved\_loan by Gender;

dump gender\_group;

count = Foreach gender\_group Generate COUNT(approved\_loan);

dump count;

--Married vs Unmarried ?

approved\_loan = Filter data by Loan\_Status == 'Y';

married\_group = Group approved\_loan by Married;

dump married\_group;

count = Foreach married\_group Generate COUNT(approved\_loan);

dump count;

--Graduate vs Not Graduate ?

approved\_loan = Filter data by Loan\_Status == 'Y';

edu\_group = Group approved\_loan by Education;

dump edu\_group;

count = Foreach edu\_group Generate COUNT(approved\_loan);

dump count;

--Self employed vs non self-employed ?

approved\_loan = Filter data by Loan\_Status == 'Y';

emp\_group = Group approved\_loan by Self\_Employed;

dump emp\_group;

count = Foreach emp\_group Generate COUNT(approved\_loan);

dump count;

--Urban vs Rural vs Semi Urban ?

approved\_loan = Filter data by Loan\_Status == 'Y';

area\_group = Group approved\_loan by Property\_Area;

dump area\_group;

count = Foreach area\_group Generate COUNT(approved\_loan);

dump count;

--Credit History vs non credit history ?

approved\_loan = Filter data by Loan\_Status == 'Y';

credit\_group = Group approved\_loan by Credit\_History;

dump credit\_group;

count = Foreach credit\_group Generate COUNT(approved\_loan);

dump count;

--Minimum applicant income and/or co-applicant income at which the any loan was approved at all ?

approved\_loan = Filter data by Loan\_Status == 'Y';

--ApplicantIncome

AppIncome = Order approved\_loan by ApplicantIncome;

low\_appIncome = LIMIT AppIncome 1;

dump low\_appIncome

--CoapplicantIncome

CoappIncome = Order approved\_loan by CoapplicantIncome;

low\_CoappIncome = LIMIT CoappIncome 1;

dump low\_CoappIncome;