New Variables for DSO562 Project 2
(Credit Cards and Cell Phone Application Data)
Note: Variables with Red Color are new-created variables.

Description of Variables	# Variables Created	Approximate CPU Time (Minute)	Names of New Columns
Day of Week of Each Application (including Monday to Sunday, 7 unique values)	1	<0.1 min	dow
Target Encoding for dow column (transferring categorical variables to numerical variables)	1	<0.1min	Dow_risk
Age of the applicant when they deliver application (treat missing values as 0)	1	<0.1m	age
Age level of the age filed (7 categories, treat missing values as "MISSING" type)	1	<0.1m	Age_level
Target Encoding for age column (transferring categorical variables to numerical variables)	1	<0.1m	Age_level_risk
Full name of applicant (including first and last name)	1	<0.1m	name
Full address of applicant (including address and 5-digit zip code)	1	<0.1m	Fulladdress
Combination of name and date of birth, address and home phone number of the applicant	3	<0.1m	Name_dob Name_fulladdress Name_homephone
Combination of full address and dob, home phone number of the applicant	2	<0.1m	Fulladdress_dob Fulladdress_homephone
Combination of home phone number and date of	2	<0.1m	Dob_homephone Homephone_name_dob

birth, name & date of birth of the applicant			
Combination of age level and address, home phone number of the applicant	2	<0.1m	Avg_level_address Avg_level_homephone
Combination of social security number and {first name, last name, address, 5-digit zip code, date of birth, home phone number, age level, name, full address, name & date of birth, age level & address, age level & home phone number}	12	<0.1m	Ssn_ {firstname, lastname, address, zip5, dob, homephone, age_level, name, fulladdress, name _dob, age_level_address, age_level_homephone}
Number of days since an application with that entity has been seen. Entities are {social security number, address, date of birth, home phone number, name, full address, name & date of birth/ full address/ home phone number, full address & date of birth/ home phone number, date of birth & home phone number/ home phone number with name, age level & address/ home phone number, social security number & first name/ last name/ address/ 5-digit zip code/ date of birth/ home phone number/ age level/ name/ full address/ name & date of birth/ age level & address/ age level & home phone number}	27	About 5 mins	{ssn, address, dob, homephone, name, fulladdress, name_dob, name_fulladdress, name_homephone, fulladdress_dob, fulladdress_homephone, dob_homephone, homephone_name_dob, age_level_address, age_level_homephone, ssn_firstname, ssn_lastname, ssn_address, ssn_zip5, ssn_dob, ssn_homephone, ssn_age_level, ssn_name, ssn_fulladdress, ssn_name_dob, ssn_age_level_address,ssn_age_level_ homephone}_day_since
Number of applications at that entity over the past n days. Entities are {social security number, address,	162	About 47 min	{ssn, address, dob, homephone, name, fulladdress, name_dob, name_fulladdress, name_homephone, fulladdress_dob,

date of birth, home phone number, name, full address, name & date of birth/ full address/ home phone number, full address & date of birth/ home phone number, date of birth & home phone number/ home phone number with name, age level & address/ home phone number, social security number & first name/ last name/ address/ 5-digit zip code/ date of birth/ home phone number/ age level/ name/ full address/ name & date of birth/ age level & address/ age level & home phone number}, n is {0,1,3,7,14,30}			fulladdress_homephone, dob_homephone, homephone_name_dob, age_level_address, age_level_homephone, ssn_firstname, ssn_lastname, ssn_address, ssn_zip5, ssn_dob, ssn_homephone, ssn_age_level, ssn_name, ssn_fulladdress, ssn_name_dob, ssn_age_level_address,ssn_age_level_homephone}_count_{0,1,3,7,14,30}
Number of applications at that entity seen in a short time window {today or past day} compared to number of applications at that entity over a longer time window {past 3,7,14,30 days}. Entities are {social security number, address, date of birth, home phone number, name, full address, name & date of birth/ full address/ home phone number, full address & date of birth/ home phone number, date of birth & home phone number/ home phone number with name, age level & address/ home phone number, social	216	About 22.58 s	{ssn, address, dob, homephone, name, fulladdress, name_dob, name_fulladdress, name_homephone, fulladdress_dob, fulladdress_homephone, dob_homephone, homephone_name_dob, age_level_address, age_level_homephone, ssn_firstname, ssn_lastname, ssn_address, ssn_zip5, ssn_dob, ssn_homephone, ssn_age_level, ssn_name, ssn_fulladdress, ssn_name_dob, ssn_age_level_address,ssn_age_level_homephone}_count_{0,1}_by_{3,7,14,30}

security number & first name/ last name/ address/ 5-digit zip code/ date of birth/ home phone number/ age level/ name/ full address/ name & date of birth/ age level & address/ age level & home phone number}			
Number of unique identy in the specific field during the past {1,3,7,14,30,60} days. Both entity and field are {social security number, full address, name & date of birth/ full address, full address & date of birth, date of birth & home phone number, age level & address/ home phone number, social security number & last name/ 5-digit zip code/ age level/ name/ full address/ name & date of birth/ age level & address/ age level & home phone number}, but with different values.	1440	About 88 mins	{ssn, fulladdress, name_dob, name_fulladdress, fulladdress_dob, dob_homephone, age_level_address, age_level_homephone, ssn_lastname, ssn_zip5, ssn_age_level, ssn_name, ssn_fulladdress, ssn_name_dob, ssn_age_level_address,ssn_age_level_ homephone}_unique_count_in_{ ssn, fulladdress, name_dob, name_fulladdress, fulladdress_dob, dob_homephone, age_level_address, age_level_homephone, ssn_lastname, ssn_zip5, ssn_age_level, ssn_name, ssn_fulladdress, ssn_name_dob, ssn_age_level_address,ssn_age_level_ homephone}_for_{1,3,7,14,30,60}
Total	1873	About 140 mins	