

**DSO 562 – FRAUD ANALYTICS**

**PROJECT 2: APPLICATIONS DATA**

**(HOMEWORK 6)**

**Submitted By: Chinmayi Bengaluru Prakash**  
**USC ID: 7142609162**

The following table provides a detailed description of variables, their description, number of variables created and approximate CPU time:

1. **Velocity and days variables (184 variables)**
2. **Relative Velocity variables (176 variables)**
3. **Entity Count Variables (unique) (2,772 variables)**
4. **New Variables – Part 1 (188 variables)**
5. **New Variables – Part 2 ( 114 variables)**

Therefore, a total of **3,434 variables** have been created (with the removal of duplicate columns while merging the data frames containing variables created at different intervals) for this project.

Name	Description of Variables	# Variables created	Approximate CPU Time
<b>VELOCITY AND DAYS SINCE VARIABLES</b>	<p><b>Velocity variables:</b> # Applications at that entity over the past n days.</p> <p><b>Days Since variables:</b> # Days since that entity has been last seen</p> <p><b>Values of n are:</b> {0, 1, 3, 7, 14, 30}</p> <p><b>Entities are:</b></p> <ol style="list-style-type: none"> <li>1. 'ssn'</li> <li>2. 'address'</li> <li>3. 'dob'</li> <li>4. 'homephone'</li> <li>5. 'name'</li> <li>6. 'fulladdress'</li> <li>7. 'name_dob'</li> <li>8. 'name_fulladdress'</li> <li>9. 'name_homephone'</li> <li>10. 'fulladdress_dob'</li> <li>11. 'fulladdress_homephone'</li> <li>12. 'dob_homephone'</li> <li>13. 'homephone_name_dob'</li> <li>14. 'ssnfirstname'</li> <li>15. 'ssnlastname'</li> <li>16. 'ssnaddress'</li> <li>17. 'ssnzip5'</li> <li>18. 'ssndob'</li> <li>19. 'ssnhomephone'</li> <li>20. 'ssnname'</li> </ol>	<b>184</b>	<b>1.327 minutes</b>

Name	Description of Variables	# Variables created	Approximate CPU Time
	21. 'ssnfulladdress' 22. 'ssnname_dob'		
<b>RELATIVE VELOCITY VARIABLES</b>	<p><b>Relative velocity variables (ratio):</b></p> <p># Applications with that entity seen in the recent past / # Applications with that same entity seen in the past {0, 1, 3, 7, 14, 30} days</p> <p><b>Entities are:</b></p> <ol style="list-style-type: none"> <li>1. 'ssn'</li> <li>2. 'address'</li> <li>3. 'dob'</li> <li>4. 'homephone'</li> <li>5. 'name'</li> <li>6. 'fulladdress'</li> <li>7. 'name_dob'</li> <li>8. 'name_fulladdress'</li> <li>9. 'name_homephone'</li> <li>10. 'fulladdress_dob'</li> <li>11. 'fulladdress_homephone'</li> <li>12. 'dob_homephone'</li> <li>13. 'homephone_name_dob'</li> <li>14. 'ssnfirstname'</li> <li>15. 'ssnlastname'</li> <li>16. 'ssnaddress'</li> <li>17. 'ssnzip5'</li> <li>18. 'ssndob'</li> <li>19. 'ssnhomephone'</li> <li>20. 'ssnname'</li> <li>21. 'ssnfulladdress'</li> <li>22. 'ssnname_dob'</li> </ol>	<b>176</b>	<b>0.742 seconds</b>
<b>ENTITY COUNT VARIABLES (UNIQUE)</b>	<p><b>Entity Count Variables (Unique):</b></p> <p># unique entities for that particular entities over the past n days:</p> <p><b>Values of n are:</b> {1, 3, 7, 14, 30, 60}</p>	<b>2772</b>	<b>63.618 mins</b>

Name	Description of Variables	# Variables created	Approximate CPU Time
	<p><b>Entities are:</b></p> <ol style="list-style-type: none"> <li>1. 'ssn'</li> <li>2. 'address'</li> <li>3. 'dob'</li> <li>4. 'homephone'</li> <li>5. 'name'</li> <li>6. 'fulladdress'</li> <li>7. 'name_dob'</li> <li>8. 'name_fulladdress'</li> <li>9. 'name_homephone'</li> <li>10. 'fulladdress_dob'</li> <li>11. 'fulladdress_homephone'</li> <li>12. 'dob_homephone'</li> <li>13. 'homephone_name_dob'</li> <li>14. 'ssnfirstname'</li> <li>15. 'ssnlastname'</li> <li>16. 'ssnaddress'</li> <li>17. 'ssnzip5'</li> <li>18. 'ssndob'</li> <li>19. 'ssnhomephone'</li> <li>20. 'ssnname'</li> <li>21. 'ssnfulladdress'</li> <li>22. 'ssnname_dob'</li> </ol>		
<p><b>NEW VARIABLES PART 1:</b></p>	<p><b>New Variables (Part 1):</b></p> <p># Applications at that entity over the past n days where:</p> <p><b>Values of n are:</b> {0, 1, 3, 7, 14, 30}</p> <p><b>Entities are:</b></p> <ol style="list-style-type: none"> <li>1. 'name_zip5'</li> <li>2. 'name_address'</li> <li>3. 'SNPD'</li> <li>4. 'ssndob_homephone'</li> <li>5. 'ssnfulladdress_homephone'</li> <li>6. 'ssnfulladdress_dob'</li> <li>7. 'ssnname_homephone'</li> <li>8. 'ssnname_fulladdress'</li> <li>9. 'dob_firstname'</li> </ol>	<p><b>188</b></p>	<p><b>0.797 mins</b></p>

Name	Description of Variables	# Variables created	Approximate CPU Time
	10. 'dob_lastname' 11. 'dob_address' 12. 'dob_zip5' 13. 'homephone_firstname' 14. 'homephone_lastname' 15. 'homephone_address' 16. 'homephone_zip5' 17. 'homephone_name_fulladdress' 18. 'homephone_fulladdress_dob'  Here, <ul style="list-style-type: none"> <li>SNPD is the concatenation of SSN, Name, Phone and DOB</li> </ul>		
<b>NEW VARIABLES PART 2:</b>	<p><b>New Variables (Part 2):</b></p> <p># Applications at that entity over the past n days where:</p> <p><b>Values of n are:</b> {0, 1, 3, 7, 14, 30}</p> <p><b>Entities are:</b></p> <ol style="list-style-type: none"> <li>'shortaddress'</li> <li>'shortadd_zip5'</li> <li>'shortadd_zip3'</li> <li>'shortdob_add'</li> <li>'shortdob_add_zip5'</li> <li>'shortdob_add_zip3'</li> <li>'ssn_zip3'</li> <li>'address_zip3'</li> <li>'dob_zip3'</li> <li>'homephone_zip3'</li> </ol> <p>Here,</p> <ul style="list-style-type: none"> <li>Entity <b>zip3</b> represents the first 3 digits of Zip5 field</li> </ul>	<b>114</b>	<b>0.418 mins</b>

Name	Description of Variables	# Variables created	Approximate CPU Time
	<ul style="list-style-type: none"> <li>Entity <b>shortaddress</b> represents the street address of the address field</li> <li>Entity <b>shortdob</b> represents the month and day (MMDD format) of dob field</li> </ul>		