

```

905 ***** 80-character banner for column width reference *****;
906 ***** 80-character banner for column width reference *****;
907 * (set window width to banner width to calibrate line length to 80 characters *;
908 ***** 80-character banner for column width reference *****;
909
910 *
911 This file prepares the dataset described below for analysis.
912 [Dataset Name] Dropouts by Race & Gender
913 [Experimental Units] California public K-12 schools
914 [Number of Observations] 58,876
915 [Number of Features] 20
916 [Data Source] The file http://dq.cde.ca.gov/dataquest/dlfile/dlfile.aspx?cLevel=
917 School&cYear=2014-15&cCat=Dropouts&cPage=filesdropouts was
918 downloaded as a text file and imported to Excel as tab-delimited text, then edited
919 to produce file project1_datasetv4.csv by setting all numeric values to number
920 types and the column labels and gender codes to text.
921 [Data Dictionary] http://www.cde.ca.gov/ds/sd/sd/fsdropouts.asp
922 [Unique ID] The columns CDS_CODE, Ethnic(ity), and Gender form a
923 composite key
924 ;
925
926 * setup environmental parameters;
927 %let inputDatasetURL =
928 https://github.com/stat6250/team-5\_project1/blob/master/project1\_datasetv4.csv?raw=true
929 ;
930
931
932 * load raw dropout dataset over the wire;
933 filename tempfile TEMP;
934 proc http
935     method="get"
936     url="&inputDatasetURL."
937     out=tempfile
938     ;
939 run;

```

NOTE: PROCEDURE HTTP used (Total process time):

```

real time          0.79 seconds
cpu time           0.04 seconds

```

```

940 proc import
941     file=tempfile
942     out=project1_raw
943     dbms=csv;
944 run;

```

NOTE: Import cancelled. Output dataset WORK.PROJECT1_RAW already exists. Specify REPLACE option to overwrite it.

NOTE: The SAS System stopped processing this step because of errors.

NOTE: PROCEDURE IMPORT used (Total process time):

```

real time          0.00 seconds
cpu time           0.00 seconds

```

```

945 filename tempfile clear;

```

NOTE: Fileref TEMPFILE has been deassigned.

```
946
947
948 * check raw dropout dataset for duplicates with respect to its composite key;

949 proc sort nodupkey data=project1_raw dupout=project1_raw_dups out=_null_;
950     by CDS_CODE Ethnic Gender;
951 run;
```

```
NOTE: There were 58875 observations read from the data set WORK.PROJECT1_RAW.
NOTE: 0 observations with duplicate key values were deleted.
NOTE: The data set WORK.PROJECT1_RAW_DUPS has 0 observations and 20 variables.
NOTE: PROCEDURE SORT used (Total process time):
    real time           0.22 seconds
    cpu time            0.04 seconds
```

```
952
953
954 * build analytic dataset from dropout dataset with the least number of columns and
955 minimal cleaning/transformation needed to address research questions in
956 corresponding data-analysis files;
957 data project1_analytic_file;
958     drop
959         year
960     ;
961
962     set project1_raw;
963 run;
```

```
NOTE: There were 58875 observations read from the data set WORK.PROJECT1_RAW.
NOTE: The data set WORK.PROJECT1_ANALYTIC_FILE has 58875 observations and 19 variables.
NOTE: DATA statement used (Total process time):
    real time           0.02 seconds
    cpu time            0.01 seconds
```