

## **Data Dictionary**

### **Object: I94Port**

<b>Field</b>	<b>Description</b>
1) I94Port	Valid and Invalid I94 port codes in SAS Dataset
2) Value	Descriptive value for I94 port code with their states

### **Object: I94Addr**

<b>Field</b>	<b>Description</b>
1) Code	Valid and Invalid State code in SAS Dataset, with invalid code being classified as other
2) Value	State Name that corresponds the abbreviated values in the Code field

### **Object: I94CityandRes**

<b>Field</b>	<b>Description</b>
1) Code	Valid and Invalid Country code in SAS Dataset
2) Value	Country Name that corresponds to I94CityandRes code field

### **Object: Airport**

<b>Field</b>	<b>Description</b>
1) Ident	Used to uniquely identify each record in Airport dataset
2) Type	Describes the type of airport (example heliport, small airport, closed airport..etc)
3) Name	Name of the Airport
4) Elevation_ft	Elevation of Airport in Feet
5) Continent	Abbreviation of Continent where the Airport is located at
6) Municipality	municipality where the airport is located at
7) Gps_code	Code which matches Ident when not null
8) Iata_code	Code which matches Ident when not null
9) Local_code	Code which matches Ident when not null
10) Coordinates	Latitude and Longitude information of the airport
11) Iso_state	Split of Iso_region into State where the airport is located at
12) Country	Split of Iso_region into country where the airport is located at

**Object: Temperature**

Field	Description
1) Year	Year in which temperature was observed
2) Month	Month in which the temperature was observed
3) Day	Day on which the temperature was recorded
4) AverageTemperature	- The Temperature that was recorded on that day in that particular location
5) AverageTemperatureUncertainty	- Denotes the level of uncertainty of the temperature, which when used along with Average Temperature, helps in accuracy in case of predictive purposes
6) City	- City where the temperature is being recorded
7) Country	- Country where the city is located at

**Object: City**

Field	Description
1) City	Name of City in United States
2) MedianAge	Median Age of all the people living in the City
3) MalePopulation	- Number of Males living in the City
4) FemalePopulation	- Number of Females living in the City
5) TotalPopulation	- Total Population of the City
6) NumberofVeterans	- Number of veterans living in the City
7) Foreignborn	- Number of residents of the city who are born in a different country
8) AverageHouseholdSize	- Describes the number of people per household
9) StateCode	- Abbreviation of states in United States
10) Race	- Categorical values that identifies which part of world people originate from
11) Count	- Number of people belonging to each race in that City

**Object: Time**

Field	Description
1) Arrdate_actualdate	- Date from SAS Dataset
2) Day	The day from Arrdate_actualdate field ( could have a value 1...31)
3) Week	The week number of that year from Arrdate_actualdate field ( could be 1..52)
4) Month	The Month number of that year from Arrdate_actualdate field (1..12)
5) Year	The year from Arrdate_actualdate field
6) Weekday	Abbreviation of Day of week from Arrdate_actualdate field (Mon... Sun)

**Object: SAS Dataset (df\_spark)**

<b>Field</b>	<b>Description</b>
1) Cicid	Unique field to describe the immigration data in SAS Dataset
2) I94yr	four digit year value denoting the year in i94 records
3) I94MON	Numeric month (1...12)
4) I94cit	Numeric value for i94cntyl which can be tied to I94CityAndResCode.txt to get the actual name
5) I94res	Numeric value for i94cntyl which can be tied to I94CityAndResCode.txt to get the actual name
6) I94port	Contains valid and invalid codes for invalid code which can be tied to I94Port.txt
7) I94mode	Takes the value 1,2,3,9 to describe Air, Sea, Land or Not reported
8) I94addr	Abbreviation of US state value which can be tied to I94ADDR.txt
9) Arrdate_Depdate -	It is the Departure Date from the USA. It has been converted from SAS date format to actual date
10) I94bir	Age of Respondent in Years
11) I94Visa	Visa codes collapsed into three categories: Business, Pleasure, Student
12) Count	Used for summary statistics
13) DTADFILE	Character Date Field - Date added to I-94 Files
14) Entdepa	Arrival Flag - admitted or paroled into the U.S.
15) Entdepd	Departure Flag - Departed, lost I-94 or is deceased
16) Matflag	Match flag - Match of arrival and departure records
17) Biryear	4 digit year of birth
18) Dtaddto	Character Date Field - Date to which admitted to U.S. (allowed to stay until)
19) Gender	Non-immigrant sex
20) Airline	Airline used to arrive in U.S.
21) Admnum	Admission Number
22) Fltno	Flight number of Airline used to arrive in U.S.
23) Visatype	Class of admission legally admitting the non-immigrant to temporarily stay in U.S.
24) Arrdate_actualdate -	is the Arrival Date in the USA. It has been converted from SAS date format to actual date
25) Arrmonth	Month from Arrdate