

Master Universitario in: "ANALISI DATI PER LA BUSINESS INTELLIGENCE"

A.A. 2013-2014

Titolo della tesi: "Tecniche avanzate di programmazione SAS e rappresentazione geografica dei dati in SAS Visual Analytics: un'applicazione pratica su open data"

Autore: Dott. Alessandro Collino

Abstract

Il presente lavoro ha come obiettivo l'illustrazione di alcune delle tecniche di programmazione avanzata e del linguaggio macro SAS, nonché delle potenzialità offerte dallo strumento web-based SAS Visual Analytics nell'esplorazione dei dati e nella loro rappresentazione su base geografica.

A tale scopo, ci siamo avvalsi dei dati reperibili sul portale Open Coesione del Dipartimento per lo Sviluppo e la Coesione Economica del Ministero dell'Economia e delle Finanze, rilasciati in modalità open data e sotto licenza Creative Commons 3.0 Italia.

Il lavoro si è articolato in due fasi principali: la prima, costituita da una serie di operazioni di tipo ETL sui dati grezzi, eseguite mediante codice e linguaggio macro SAS in ambiente SAS Foundation 9.4; la seconda, relativa all'esplorazione e alla localizzazione dei dati mediante SAS Visual Analytics 6.4 su LASR Analytic Server.

La prima parte del progetto è descritta nel dettaglio nei primi cinque capitoli, dei quali il più importante è il terzo, quello in cui si mettono a confronto cinque differenti tecniche per il merging tra dataset, che ci hanno consentito di introdurre ed illustrare funzioni, oggetti e istruzioni avanzate del linguaggio SAS. Sia nel primo che nel terzo capitolo viene fatto inoltre ampio ricorso al linguaggio macro SAS.

La seconda parte del lavoro è illustrata invece nel sesto ed ultimo capitolo, dedicato al Visual Analytics: la basetable prodotta nella prima parte del progetto è stata prima importata su LASR Server (caricamento in-memory), quindi si è passati all'esplorazione dei dati, giungendo infine alla creazione di un Report interattivo accessibile via URL in un browser, mediante il quale visualizzare informazioni relative agli importi finanziati o ai costi sostenuti aggregate a livello nazionale, regionale o provinciale, filtrando di volta in volta, in tempo reale, sulla base dell'anno di riferimento, dei settori di interesse o della localizzazione geografica dei progetti.