

Software Visuali per Analisi Avanzate

Corso sostitutivo di Tirocinio



Dott. Valerio Morfino – Sistemi Visuali per Analisi Avanzate a.a. 2019-2020

1

Calendario del
corso v.2.0



Giorno	Durata	Argomenti
Martedì 5/5 16.00 – 18.00	2 ore	Presentazione del Corso Dati Strutturati, semi-strutturati e non strutturati Software da installare per il corso
Sabato 9/5 9.30 - 13.30	4 ore	Database relazionali MySQL Database, Tabelle, Righe, Colonne Linguaggio SQL: Introduzione, Select, Where,
Martedì 12/5 16.00 – 18.00	2 ore	Linguaggio SQL: Inner Join, Left Join, Right Join, Order By
Sabato 16/5 9.30 - 13.30	4 ore	Linguaggio SQL: Query di aggregazione Cenni su Indici Cenni sulle DDL Esercitazione finale SQL
Martedì 19/5	2 ore	Programmazione con Sas: Introduzione a SAS Studio Struttura di un programma SAS
Venerdì 22/5	4 ore	Programmazione SAS – Ospite docente di SAS Institute
Sabato 30/5	4 ore	Programmazione con Sas SQL con SAS Esercitazione finale SAS
Venerdì 5/6	2 ore	Incontro con HR Manager L'orizzonte più lungo dell'advanced analytics: il Deep Learning

Dott. Valerio Morfino – Software Visuali per Analisi Avanzate a.a. 2019-2020

2

Comunicazioni di servizio

- Link Slide del corso
 - <https://github.com/valeriovvv/Corso-Software-Visuali-per-Analisi-Avanzate/tree/master/Modulo-2>
- Rilevamento presenze: Quando richiesto, inserire in chat: Nome, Cognome e Matricola
- Consenso per gli attestati di SAS
- Consegna elaborati SQL + SAS
- Spostamento ultima lezione a Giovedì 4/6 ore 16.30 – 18.30

Attestazione di partecipazione di SaS

- Pubblicherò sulla chat la seguente formula:
«Digitando nome, cognome ed indirizzo email autorizzo il trattamento dei miei dati personali al fine di predisporre un attestato di partecipazione per la lezione tenuta da SaS il 22/05/2020 nell'ambito del corso sostitutivo di tirocinio Software Visuali per Analisi Avanzate presso il dipartimento DEMM dell'Università del Sannio»
- Se accettate, digitate nome, cognome ed indirizzo email istituzionale

Sommario

SAS programming

Utilizzo della documentazione

Installazione Driver Mysql per Tableau

Esercitazione Finale

Utilizzo della documentazione



Documentazione di riferimento

documentation.sas.com/

(selezionare la tecnologia SAS prima di eseguire la ricerca)



Problemi e casi specifici:

stackoverflow.com/



Ricerca globale:

www.google.it/

(siti in genere utili: es. tutotialpoint.com)



Esempi di ricerca

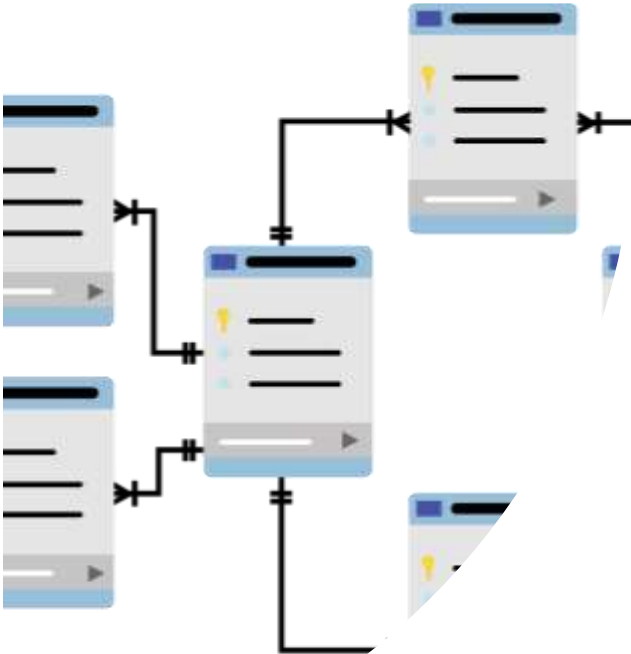
- Su documentation.sas.com
 - Andare nella sezione Documentazione Generica / Programmazione in SAS
 - Cercare «donut»
- Su Google:
 - Cercare: «Sas donut plot»

7

Installazione Driver Mysql con Tableau

- Avviare Mysql (passo non necessario se si una db4free)
- Aprire Tableau
- Cliccare sul link al driver
- Scaricare Driver Mysql
- Riavviare Tableau
- Connettersi

8



Esercitazione finale

Elaborato finale Tableau

- Il Coronavirus (COVID-19) è una malattia infettiva causata da un coronavirus recentemente scoperto.
- Il modo migliore per prevenire e rallentare la trasmissione è **essere ben informato sul virus COVID-19**, sulla malattia che provoca e su come si diffonde. Proteggi te stesso e gli altri dalle infezioni lavandoti le mani o usando un disinfettante a base di alcol frequentemente e senza toccarti il viso.

[World Health Organization
<https://www.who.int/health-topics/coronavirus>]



Dashboard della protezione civile

- La protezione civile ha messo a disposizione una dashboard per il monitoraggio, aggiornata giornalmente:
<http://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html#/b0c68bce2cce478eaac82fe38d4138b1>
- I dati sono anche disponibili in modalità opensource su GitHub:
<https://github.com/pcm-dpc/COVID-19>



11

Analisi Covid-19 con SAS ed SQL

- Scaricare i dati della protezione civile più aggiornati comprendenti i dati per provincia
- Scaricare un dataset con i dati sugli abitanti per provincia in Italia
- Analizzare i dataset sia singolarmente che in maniera correlata

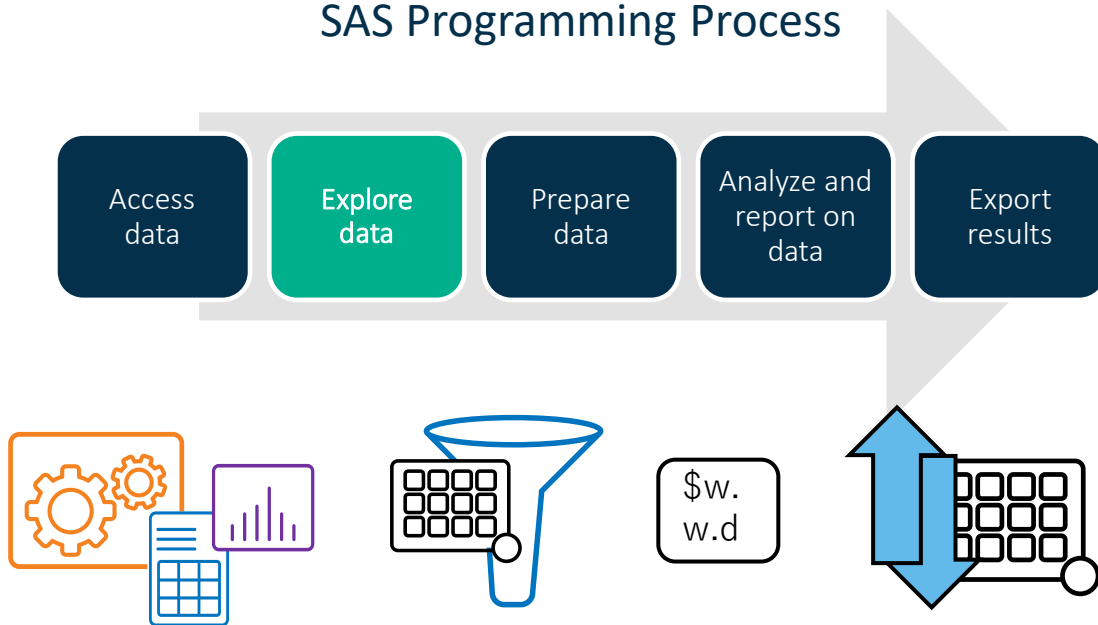


12

Dettagli e tips

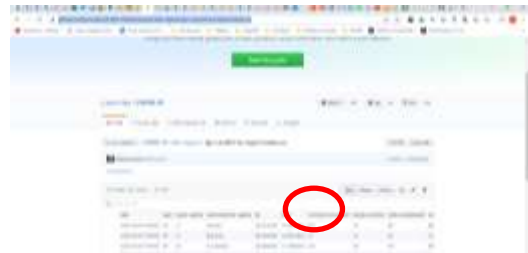
- Scaricare i dataset dai link:
 - <https://github.com/pcm-dpc/COVID-19/blob/master/dati-province/dpc-covid19-ita-province.csv>
 - <https://github.com/valeriovv/Corso-Software-Visuali-per-Analisi-Avanzate/blob/master/Modulo-2/PopolazionePErProvincia.csv>
- Dovranno essere consegnati:
 - Programma SAS
 - Report e grafici in un unico file in un formato tra: PDF, PPT, RTF generato da SAS, con titoli, footnote e formattazioni
 - Eventuali output in csv o xlsx
- Procedere seguendo il SAS programming process

SAS Programming Process



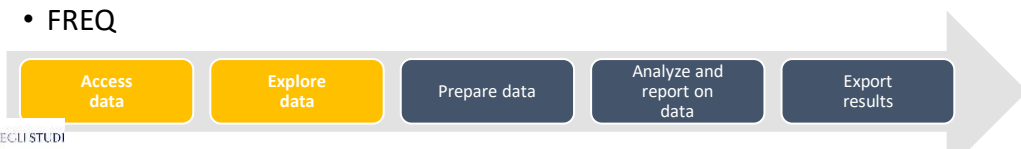
Scaricare i dati

- Aprire il link: <https://github.com/pcm-dpc/COVID-19/blob/master/dati-province/dpc-covid19-ita-province.csv>
- Cliccare su RAW
- Salvare la pagina che si apre con il nome: dpc-covid19-ita-province.csv
- Eseguire stessa procedura per il file: <https://github.com/valeriiovvv/Corso-Software-Visuali-per-Analisi-Avanzate/blob/master/Modulo-2/PopolazionePERProvincia.csv>



Analisi dati Covid-19 - 1

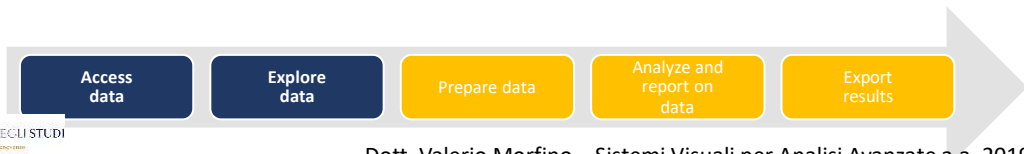
- **Accedere ai dati:**
 - Scaricare i file con i dataset, caricarli su SAS ed importarli come tabella
 - Importare le tabelle in una libreria chiamata COVID19
- **Esplorare i dati:**
 - Esplorare singolarmente le due tabelle utilizzando le proc:
 - PRINT (limitarsi a 10 osservazioni ed alle colonne più significative)
 - MEANS
 - UNIVARIATE (concentrarsi su Extreme observations table e su Missing values)
 - FREQ



Analisi dati Covid-19 – 2

• Analisi dati Covid-19 per Provincia e per Regione:

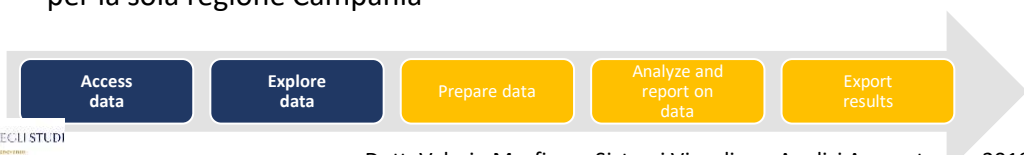
- Creare un report con due campi: Provincia e Contagiati per Provincia
 - Suggerimento: Il dataset della protezione civile propone delle serie temporali con il campo dei contagi valorizzato come totale incrementale (running total), quindi basta prendere il valore massimo.
- Creare un report con due campi: Regione e Contagiati per Regione
 - Suggerimento: E' necessario sommare i valori di tutte le province di ogni regione
- Creare un Pie Chart o un Donut chart con i contagi per Regione
- Esportare i dati dei contagiati per Provincia in csv o in xlsx



Analisi dati Covid-19 - 3

• Analisi dati Covid-19 per Provincia e per Regione:

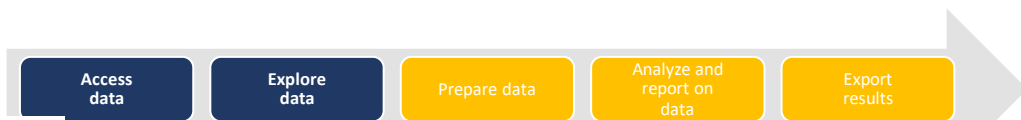
- Collegare i due dataset creando un'unica tabella con i seguenti campi: Regione, Provincia, Contagi, Residenti, Superficie Km_q, Tasso_Contagi_Residenti, Tasso_Contagi_kmq. La tabella dovrà contenere il valore più recente dei contagiati (cioè il massimo).
 - Suggerimento: Join sql, campi calcolati per i due tassi
- Creare un Report che mostri tutte le colonne con i valori ordinati per numero di contagi decrescente
- Creare uno Scatter Diagram che mostri: Popolazione, Contagiati e Provincia per le sole province con più di 3000 Contagiati
- Creare uno Scatter Diagram che mostri: Popolazione, Contagiati e Provincia per la sola regione Campania



Analisi dati Covid-19 - 4 (opzionale)

• Analisi dati Covid-19 per Regione:

- Creare un Report che mostri: Regione, Contagi, Residenti_Regione, Superficie Km_q_regione, Tasso_Contagi_Residenti, Tasso_Contagi_kmq. Valori ordinati per numero di contagi decrescente
- Creare uno Scatter Diagram che mostri: Popolazione, Contagiati e Regione per le sole Regioni con più di 4000 Contagiati
- Creare uno Scatter Diagram che mostri: Kmq, Contagiati e Regione per le sole Regioni con più di 4000 Contagiati
- Utilizzando il Wizard, creare una Mappa a Bolle contenente i contagi per Provincia



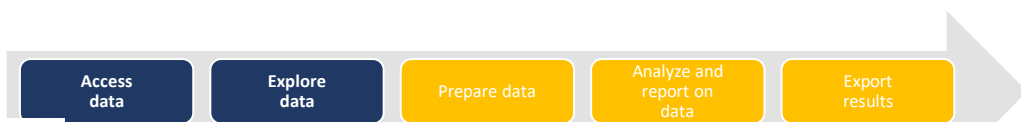
Analisi dati Covid-19 - 5 (opzionale)

• Analisi dati Covid-19:

- Aprire il menù Processi e Utilità
- Utilizzare i Wizard presenti per analizzare ulteriormente i dati del Dataset

• Alcuni spunti:

- Grafici
- Mappe
- Machine Learning non supervisionato (Clustering)
- Analisi di serie temporali
- Analisi delle correlazioni



Tips

- option validvarname=v7;
- Usare proc print e le sue opzioni per formattare i report
- Usare import wizard per i csv
- Utilizzare procexport per esportare i dati (oppure il wizard)
- Utilizzare i commenti per rendere il codice comprensibile
- Utilizzare la formattazione (anche automatica)
- Utilizzare le direttive title, footnote, footnote2, footnote3 all'interno delle proc per generare i diagrammi per inserire delle descrizioni

Grazie per l'attenzione



<https://it.linkedin.com/in/valerio-morfino>



vmorfino@unisannio.it