

Software Visuali per Analisi Avanzate

Corso sostitutivo di Tirocinio



UNIVERSITÀ DEGLI STUDI
DEL SANNIO Benevento
DEMM
DIPARTIMENTO DI DIRITTO, ECONOMIA,
MANAGEMENT E METODI QUANTITATIVI

Dott. Ing. Valerio Morfino – Sistemi Visuali per Analisi Avanzate a.a. 2020-2021

1

Calendario Modulo 2

Giorno	Durata	Argomenti
Venerdì 30/04	2	Presentazione del Corso Dati Strutturati, semi-strutturati e non strutturati Software da installare per il corso
Giovedì 6/5	4	Database relazionali MySQL Database, Tabelle, Righe, Colonne Linguaggio SQL: Introduzione, Select, Where,
Venerdì 7/5	2	I sistemi informativi aziendali - Ospite
Giovedì 13/5	4	Linguaggio SQL: Inner Join, Left Join, Right Join, Order By Linguaggio SQL: Query di aggregazione Cenni su Indici Cenni sulle DDL
Venerdì 14/5	2	Esercitazione Finale SQL
Giovedì 20/5	4	Introduzione a SAS Studio Struttura di un programma SAS Programmazione SAS
Sabato 22/5	2	Programmazione con Sas
Sabato 29/5	4	SAS Esercitazione finale SAS Ospite



UNIVERSITÀ DEGLI STUDI
DEL SANNIO Benevento
DEMM
DIPARTIMENTO DI DIRITTO, ECONOMIA,
MANAGEMENT E METODI QUANTITATIVI

Dott. Ing. Valerio Morfino – Sistemi Visuali per Analisi Avanzate a.a. 2020-2021

2

Comunicazioni di servizio

- Link Slide del corso
 - <https://github.com/valerio75/Software-Visuali-per-Analisi-Avanzate-2020-2021>
- Consegna Elaborati Modulo 2 - 10 Giugno
- Scaricare il file di esempio:
 - <https://github.com/valerio75/Software-Visuali-per-Analisi-Avanzate-2020-2021/blob/main/TitoloEdExport.sas>

Sommario

SAS programming

Utilizzo della documentazione

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)

Dott. Ing. Valerio Morfino – Sistemi Visuali per Analisi Avanzate a.a. 2020-2021

5

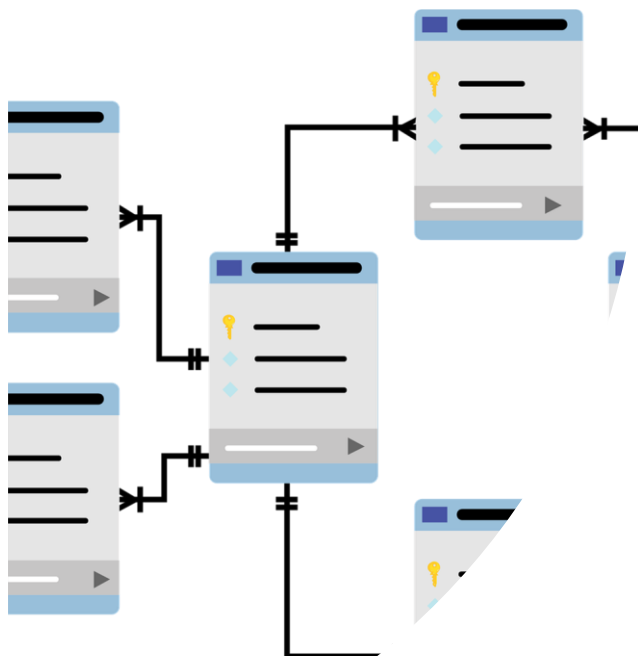


Esempi di ricerca

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

Dott. Ing. Valerio Morfino – Sistemi Visuali per Analisi Avanzate a.a. 2020-2021

6



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>]





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

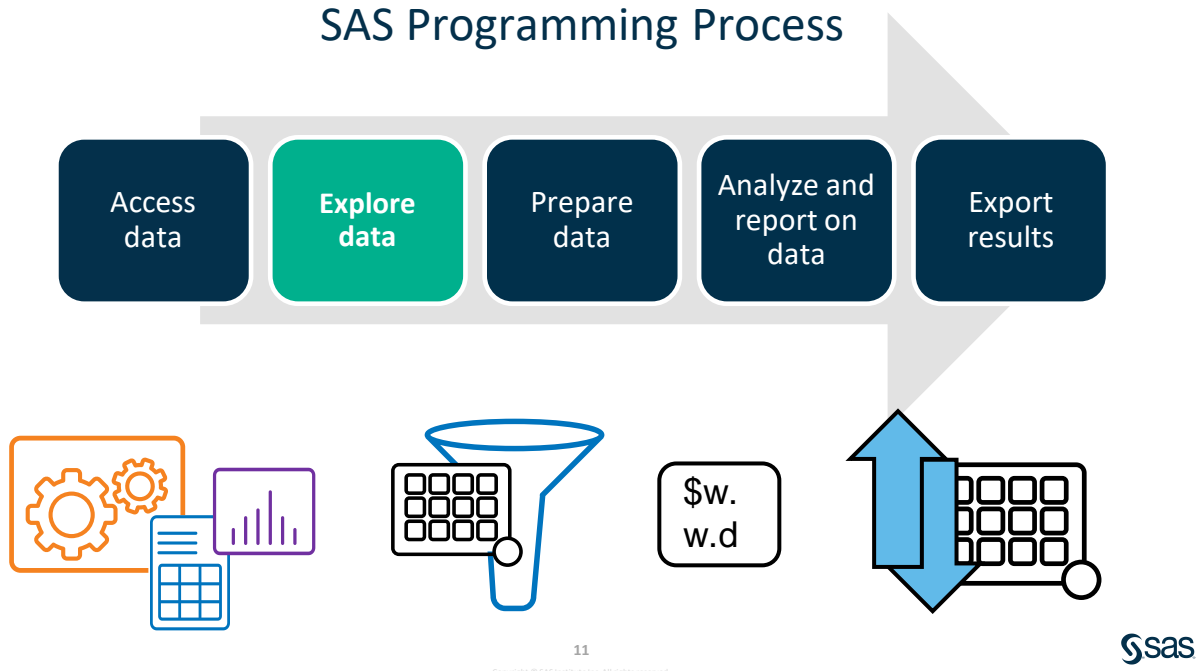
9

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/valeriovvv/Corso-Software-Visuali-per-Analisi-Avanzate/blob/master/Modulo-2/PopolazionePErProvincia.csv>
- Procedere seguendo il SAS programming process

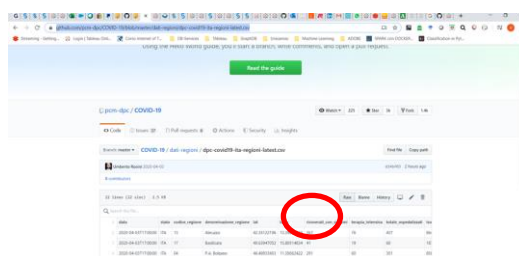
10

SAS Programming Process



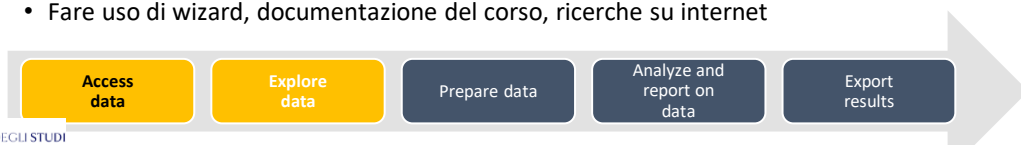
Scaricare i dati

- 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/valeriovv/Corso-Software-Visuali-per-Analisi-Avanzate/blob/master/Modulo-2/PopolazionePERProvincia.csv>



Analisi dati Covid-19 - Deliverables

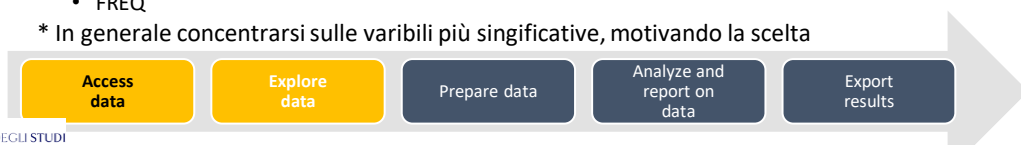
- Dovranno essere consegnati:
 - Programma SAS (codice sorgente)
 - Report e grafici in un unico file in un formato a scelta tra: PDF, PPT, RTF generato da SAS, con titoli, footnote ed eventuali formattazioni
 - Non vanno consegnati gli output in csv o xlsx
- Per ognuno degli output richiesti:
 - Inserire l'output richiesto (es. Report, Grafico) con Titolo e Footnote che contengano descrizioni adeguate
 - Nel codice sorgente inserire commenti, ove necessario, per rendere il codice più leggibile
 - Formattare il codice
 - Fare uso di wizard, documentazione del corso, ricerche su internet



Analisi dati Covid-19 - 1

- **Accedere ai dati:**
 - Creare su SAS una cartella denominata COVID19
 - Scaricare i file con i dataset, caricarli su SAS nella cartella COVID-19
 - Importare le tabelle in una libreria chiamata COVID19
 - Creare un Dataset con i contagi che contenga i dati a partire dal 01/04/2021
 - Esportare i risultati di tutte le analisi con il comando ODS
- **Esplorare i dati:**
 - Esplorare singolarmente le due tabelle utilizzando le proc *:
 - CONTENTS
 - PRINT (limitarsi a 10 osservazioni ed alle colonne più significative)
 - MEANS
 - UNIVARIATE (concentrarsi su Extreme observations table e su Missing values)
 - FREQ

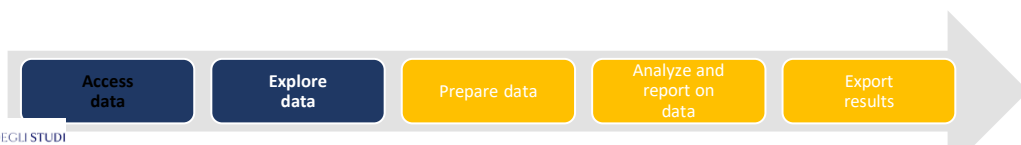
* In generale concentrarsi sulle variabili più significative, motivando la scelta



Analisi dati Covid-19 – 2

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

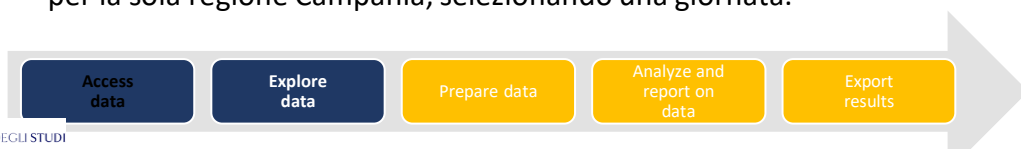
- Creare un report con due campi: Provincia e Contagiati per Provincia per una specifica giornata presente nel dataset ordinati per contagi
- Creare un report con due campi: Regione e Contagiati per Regione
 - Suggerimento: E' necessario sommare i valori di tutte le province di ogni regione per la stessa data (da scegliere)
- Creare un Pie Chart o un Donut chart con i contagi per Regione in un determinato giorno
- Esportare i dati dei contagiati per Provincia in csv o in xlsx utilizzando la procedura guidata di export oppure la funzione (vedere documentazione)



Analisi dati Covid-19 - 3

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

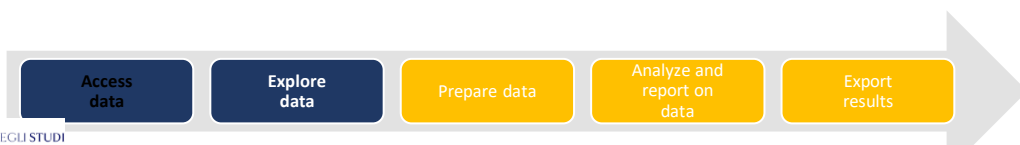
- Collegare i due dataset creando un'unica tabella con i seguenti campi: Regione, Provincia, Contagi, Residenti, Superficie Km_q, Tasso_Contagi_Residenti, Tasso_Contagi_kmq. Filtrare per una specifica giornata i contagi.
 - Suggerimento: Join sql, campi calcolati per i due tassi
- Creare un Report che mostri tutte le colonne con i valori ordinati per numero di contagi decrescente. Filtrare per una specifica giornata.
- Creare uno Scatter Diagram che mostri: Popolazione, Contagiati e Provincia per le sole province con più di 5000 Contagiati, selezionando una giornata.
- Creare uno Scatter Diagram che mostri: Popolazione, Contagiati e Provincia per la sola regione Campania, selezionando una giornata.



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. Filtrare per una specifica giornata.
- Creare uno Scatter Diagram che mostri: Popolazione, Contagiati e Regione per le sole Regioni con più di 5000 Contagiati
- Creare uno Scatter Diagram che mostri: Kmq, Contagiati e Regione per le sole Regioni con più di 5000 Contagiati
- Utilizzando il Wizard, creare una Mappa a Bolle contenente i contagi per Provincia



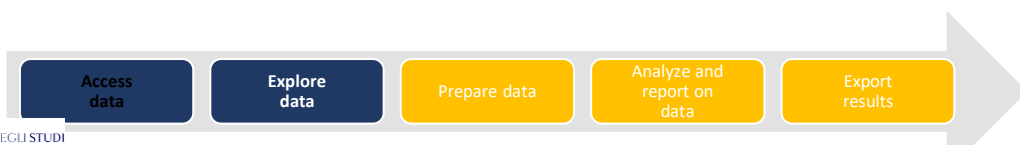
Analisi dati Covid-19 - 5 (opzionale)

• 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

Tips

- Per filtrare per data utilizzare nella condizione where la seguente sintassi: where datepart(data) = '<DATA>'d
Ad esempio se si ha una query SQL:
select * from mia_tabella where datepart(data) = '01MAY21'd
- Utilizzare ods per la generazione dell'output

Grazie per l'attenzione



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



vmorfino@unisannio.it