



UNIVERSITA' POLITECNICA DELLE MARCHE

**Corso di Laurea Magistrale in Ingegneria Informatica e
dell'Automazione**

**CORSO DI DATA SCIENCE:
PROGETTO BUSINESS INTELLIGENCE**

Caprari David

Cingolani Cristian

Anno Accademico 2021-2022

Sommario

Introduzione.....	3
Dataset	4
Qlik Sense.....	5
Introduzione allo strumento.....	5
Inserimento dataset e ETL.....	5
Costruzione ed analisi dei fogli	7
Economy.....	8
Fertility	10
Unemployment	11
Education.....	12
Italy.....	13
Tableau Desktop.....	14
Introduzione allo strumento.....	14
Inserimento dataset e preparazione.....	14
Costruzione fogli di lavoro	16
Allestimento dashboard	16
Dashboard World.....	17
Dashboard Fertility	18
Dashboard Unemployment	19
Dashboard Italy	21
Power BI.....	23
Introduzione allo strumento.....	23
Connessione Origine Dati.....	23
Elaborazione Dati	23
Configurazione Modello Dati.....	25
Creazione Visualizzazioni	26
GDP	26
Fertility	27
Youth Population.....	27
Un/Employment by Age Group & instruction Level	29
Italy.....	30
Considerazioni sugli strumenti utilizzati	31

Introduzione

Il progetto, di cui questa relazione riporta alcune fasi di realizzazione, ha come obiettivo quello di un'analisi di un dataset al fine di sfruttare la potenza di diversi software per la business intelligence in modo da ottenere una prima astrazione di informazioni che, all'interno del caso didattico in oggetto, possano avere valore in un ipotetico contesto aziendale.

In questa relazione verrà quindi inizialmente descritto il dataset scelto, verranno poi riportate le diverse fasi di progetto, ognuna delle quali contraddistinta dall'utilizzo di diversi software o tecniche. Si avrà quindi cura di riportare le fasi di utilizzo di ogni software e l'estrazione di informazioni che questo rende possibile sfruttando le capacità grafiche di questi ultimi e raccontando quanto ottenuto.

I software utilizzati per l'analisi di business intelligence giacciono tutti nel quadrante "Leaders" dei Magic Quadrants dell'azienda di IT consulting Gartner. Le aziende QlikTech, Tableau Software e Microsoft che producono rispettivamente QlikSense, Tableau Desktop e Power BI sono al vertice per l'analisi e la rappresentazione dei dati in ambito aziendale.

Dataset

Il dataset scelto per lo svolgimento del progetto è stato reperito attraverso la piattaforma Kaggle che mette a disposizione una grande quantità di raccolte dati a scopi di ricerca ed uso personale.

Il dataset, all'interno della piattaforma prende il nome di County Economic Indicators – OECD (<https://www.kaggle.com/datasets/michalisz/country-economic-indicators-oecd>).

Riporta 17 file in formato .csv ognuno con diversi parametri di sviluppo economico, e non solo, di molti degli stati del globo. Ognuno di questi riporta al suo interno otto colonne che definiscono la provenienza geografica del dato, l'anno a cui esso fa riferimento e diversi valori riguardanti categorie di sviluppo economico quali il GDP (equivalente dell'italiano Prodotto Interno Lordo), disoccupazione, tasso di fertilità, debito governativo e altri.

L'aspetto temporale sarà di fondamentale importanza per diverse analisi, in particolar modo per eseguire predizioni.

La scelta fatta dal gruppo di progetto è quella di utilizzare un certo numero di questi file .csv e di non catturarne la totalità per non appesantire troppo il carico di lavoro.

I file utilizzati sono:

- *GDP.csv*
- *Average_wage.csv*
- *Employment_rate_by_age_group.csv*
- *Unemployment_rate_by_education_level.csv*
- *Long_temp_unemployment_rate.csv*
- *Youth_unemployment_rate.csv*
- *Hours_worked.csv*
- *Average_wage.csv*
- *Young_population.csv*

Alcuni file come GDP.csv presentano al loro interno dei campi contenenti più misure di interesse (es: nel campo Measure di GDP.csv compare sia la voce USD_CAP che la voce MLN_USD). Per gestire questo aspetto sono state utilizzate varie tecniche in funzione delle caratteristiche di ciascun software: split in due tabelle per Qlik, filtraggio in fase di visualizzazione per Tableau e operazioni di pivot per PowerBi.

Per ciascun file, inoltre, sono state eseguite in tutti i software operazioni di ETL quali filtraggio, parsing, eliminazione colonne, ridenominazione colonne, ecc...

Oltre ai file precedentemente indicati è stato integrato un ulteriore file, *continents.csv*, in cui ad ogni nazione sono associati vari standard di abbreviazione ed il continente di appartenenza. Questo, come vedremo, sarà utile per ovviare ad un mancato riconoscimento delle nazioni da parte dei software o per eseguire confronti tra i vari continenti.

Qlik Sense

Introduzione allo strumento

La suite Qlik offre software di visualizzazione dei dati e di business intelligence che permette il rapido sviluppo di fogli completamente personalizzabili affinché si possano offrire rapidamente informazioni utili sui dati a disposizione.

Tra i tool utilizzati in questo progetto, Qlik Sense è quello che si pone più in basso all'interno della rappresentazione data dai Magic Quadrants di Gartner, questo a significare che lo strumento viene inteso come meno capace di riportare una piena attualità e utilità di prodotto, meno capace quindi di un completo orientamento ai mercati ed al loro cambiamento.

Risulta quindi essere un ottimo strumento per la rappresentazione grafica dei dati ed ha il pregio di essere molto grafico e di semplice utilizzo, accompagnando, anche con suggerimenti molto mirati, l'utilizzatore alla generazione del suo rapporto.

Al fine del progetto, viene utilizzato Qlik Sense nella sua ormai più diffusa versione all'interno della piattaforma Qlik Cloud, delegando quindi l'elaborazione ed il caricamento dei dati alla sua struttura cloud, comunicando con l'utente attraverso browser.

Inserimento dataset e ETL

L'inserimento dei dati in Qlik risulta piuttosto semplice ed è piuttosto intuitivo. Il tutto viene accompagnato da rappresentazioni tabellari e rappresentazioni a bolla che aiutano a comprendere le relazioni che si vengono a creare tra le varie tabelle.

Come precedentemente indicato, in Qlik, sono state splittate alcune tabelle per rendere la fase di analisi e visualizzazione più semplice e intuitiva. In particolare, sono stati splittati i seguenti file:

- GDP in GDP_MLN e GDP_CAP (corrispondenti al PIL e al PIL pro capite)
- Employment_rate_by_age_group in EMPLOYMENT_RATE_BY_AGE_GROUP_15-25 e ..._25-74
- Unemployment_rate_by_education_level in UNEMPLOYMENT_RATE_BY_EL_BUPPSRY, ..._UPPSTRY_NTRY, ..._TRY
- Youth_unemployment_rate in Y_UNEMPLOYMENT_MEN E Y_UNEMPLOYMENT_WOMEN

Sono state dunque create per ciascuna tabella delle Key composte da Location+Time in modo da poter metterle in relazione.

In alcune tabelle è stato svolto un filtraggio che ha permesso di eliminare alcune righe che seguivano una distribuzione temporale non annuale ma ad esempio per quartile.

Sono state eliminate anche tutte quelle istanze che rappresentavano gruppi di nazioni in quanto non erano utili al fine dell'analisi e creavano problemi in fase di visualizzazione. Alcuni esempi: mkd, ea19, eu27_2020, oecd, g7, oecde...

Un esempio del risultato ottenuto:

GDP_MLN
GDP.csv
Colonne: 8
Righe: 2498
Trasforma tramite UnPivot
Aggiungi campo
Selezione dati da sorgente
X

GDP_MLN.GDP.LOCATI...	GDP_MLN.GDP.INDICA...	GDP_MLN.GDP.SUBJECT	GDP_MLN.GDP.MEASURE	GDP_MLN.GDP.FREQU...	GDP_MLN.GDPTIME	GDP_MLN.GDPValue	GDP_MLN.KEY
ALB	GDP	TOT	MLN_USD	A	1997	8555.927	ALB1997
ALB	GDP	TOT	MLN_USD	A	1998	9451.507	ALB1998
ALB	GDP	TOT	MLN_USD	A	1999	10792.634	ALB1999
ALB	GDP	TOT	MLN_USD	A	2000	11926.2	ALB2000
ALB	GDP	TOT	MLN_USD	A	2001	13157.36	ALB2001
ALB	GDP	TOT	MLN_USD	A	2002	14221.925	ALB2002
ALB	GDP	TOT	MLN_USD	A	2003	15198.213	ALB2003
ALB	GDP	TOT	MLN_USD	A	2004	16429.396	ALB2004
ALB	GDP	TOT	MLN_USD	A	2005	17663.352	ALB2005
ALB	GDP	TOT	MLN_USD	A	2006	19650.703	ALB2006
ALB	GDP	TOT	MLN_USD	A	2007	21631.653	ALB2007
ALB	GDP	TOT	MLN_USD	A	2008	24251.628	ALB2008
ALB	GDP	TOT	MLN_USD	A	2009	25800.372	ALB2009
ALB	GDP	TOT	MLN_USD	A	2010	28048.719	ALB2010
ALB	GDP	TOT	MLN_USD	A	2011	29655.634	ALB2011
ALB	GDP	TOT	MLN_USD	A	2012	30530.351	ALB2012

GDP_MLN.GDP.LOCATION Riepilogo (Dimensione)

Valori distinti: 58 Valori totali: 2348

Anteprima distribuzione valori distinti

Valori null

Valori null: 150 / 2498

Sostituisci Imposta null Ordinamento Dividi

Selezionare i valori da sostituire. È possibile utilizzare fino a 99 valori di sostituzione.

Valori distinti

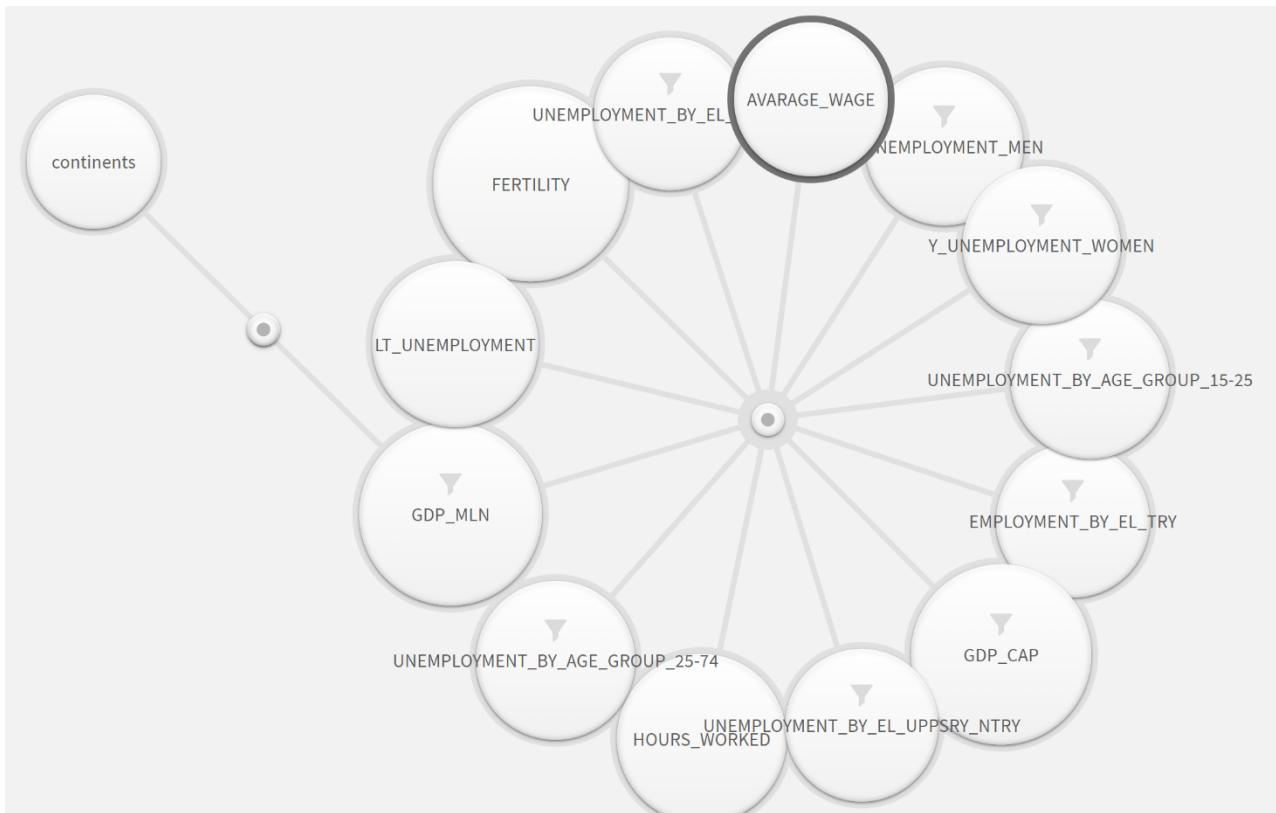
- (Null)
- AUS
- AUT
- BEL
- CAN
- CZE
- DNK
- ETN

Selezionare uno o più valori da sostituire con un nuovo valore
È possibile sostituire più insiemi di valori

Annulla Sostituisci

Infine, per ottenere delle aggregazioni per continente è stata collegata la tabella continents.

Le relazioni fra tabelle possono essere osservate nella figura di seguito:



Costruzione ed analisi dei fogli

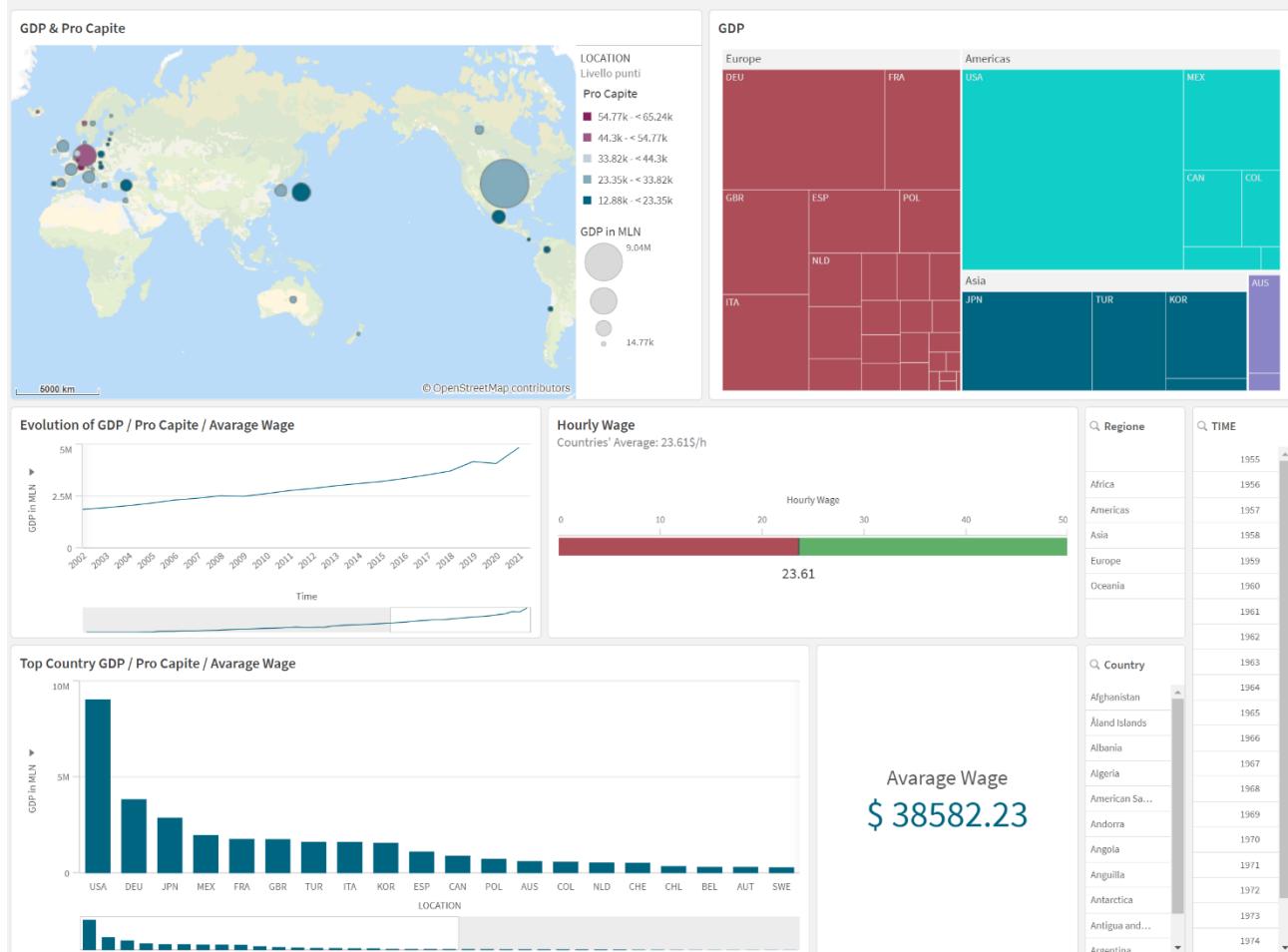
Per poter distinguere cinque macro-argomenti di interesse sono stati creati altrettanti fogli di lavoro:

- Economy: riguardante gli aspetti economici delle varie nazioni ed i trend di tali variabili economiche
- Fertility: vengono confrontate le nascite delle varie aree del mondo, viene analizzato il trend e le possibili correlazioni con altre variabili
- Unemployment: vengono confrontati i valori di disoccupazione delle varie aree del mondo, viene analizzato il trend e le possibili correlazioni con altre variabili, viene confrontata la disoccupazione fra uomini e donne
- Education Level: viene correlato il livello di disoccupazione relativo ai vari gradi di istruzione con altre variabili, viene visto l'andamento temporale e spaziale di tale parametro
- Italy: viene analizzato l'andamento temporale di alcune variabili viste negli altri fogli ma con un focus sull'Italia

PS. Nelle seguenti sezioni vengono spiegati puntualmente i singoli fogli ma è stato deciso di non visualizzare immagini riguardanti i singoli anni/stati/continenti per non appesantire la presentazione del progetto con casi troppo specifici.

Economy

Economy



Il primo foglio che prendiamo in esame riguarda l'aspetto economico.

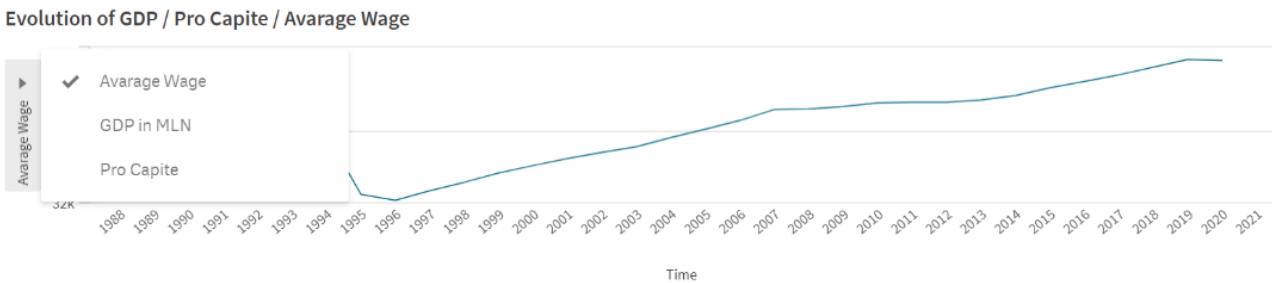
Innanzitutto, in basso a destra sono stati inseriti dei filtri che permettono di eseguire delle analisi sui singoli paesi/continenti ed in determinati anni, è sufficiente un click nel campo di interesse.

Il primo grafico che incontriamo in alto a sinistra è una mappa con cerchi di grandezza e colore differente. Come si può osservare dalla legenda la grandezza è proporzionale al GDP ovvero al PIL del paese, mentre il colore è correlato al PIL pro capite. Per le motivazioni esposte precedentemente, i valori che vediamo in questo grafico ed in molti altri sono una media dei valori nel tempo (o in altri casi una media/somma tra le varie nazioni). Appare comunque evidente come storicamente USA e Germania hanno i valori di PIL più alti fra i vari stati. Se parliamo di PIL pro capite, invece, spiccano paesi come la Svizzera e Lussemburgo. Andando a selezionare gli anni più recenti la situazione rimane simile ma non invariata per lo sviluppo economico di alcuni paesi. Non compare la Cina nella mappa ma sicuramente rappresenta una delle principali economie globali.

Nella mappa ad albero nella parte destra dell'immagine è ancora più evidente quanto detto in precedenza, ma appare evidente come i continenti storicamente più potenti dal punto di vista economico sono l'America, l'Europa e l'Asia(tenendo conto della grande assente, ovvero della Cina). L'Africa appare molto debole economicamente sia per valori di PIL molto bassi che per assenza di dati storici, soprattutto nei decenni passati.

Nel grafico lineare a sinistra si può osservare come il GDP globale sia mediamente cresciuto nel tempo, questo potrebbe essere dovuto ad una forte crescita demografica (soprattutto in Asia e Africa) ed al miglioramento della tecnologia, oltre che per innumerevoli altre cause.

È possibile selezionare all'interno dello stesso grafico una fra le seguenti variabili: GDP, GDP pro capite e Avarage Wage. Di seguito una immagine dimostrativa.



Ad un aumento del GDP corrisponde anche un aumento nel tempo del GDP pro capite e ad un aumento del salario medio, quindi si ha una qualità della vita ed una produttività migliore.

Il misuratore che vediamo a destra del grafico appena descritto è stato ottenuto dividendo il salario medio annuale per il numero di ore lavorate in un anno. È stato aggiunto un valore di riferimento che corrisponde alla media globale di circa 23.61\$/h. Se analizziamo lo storico medio dell'Italia otteniamo il seguente grafico.



Possiamo notare, dunque, come il salario orario in Italia è inferiore alla media globale. Da sottolineare che l'assenza di molti paesi, soprattutto africani, porti ad una sovrastima del salario storico medio globale.

Nel grafico a barre in basso a sinistra vi è un ulteriore conferma dei paesi con i valori di GDP più alti. Anche questa volta il selettore permette di selezionare anche il GDP pro capite ed il salario medio. L'Italia non spicca particolarmente proprio su quest'ultimo aspetto.

Infine, visualizziamo il salario annuale medio in una KPI e notiamo che il valore medio corrisponde a 38582\$. In questo caso selezionando l'Italia si ottiene un valore leggermente superiore alla media globale.

Fertility



Il foglio che andremo ad analizzare ora prova ad indagare correlazioni fertility rate ed altre variabili, oltre ad analizzare il trend e la differenza tra le varie aree del globo.

I primi due grafici sono scatterplot e sono interessanti in quanto, contrariamente a quanto si possa pensare, non vi è una correlazione così marcata tra fertility rate e GDP (quindi la potenza economica totale di una nazione) o tra fertility rate e la disoccupazione nella fascia di età 25-75. Viene supposto in questo grafico e nei successivi che il valore di soglia tra una crescita ed una decrescita demografica sia 2. In realtà nel primo grafico si può scorgere una leggera relazione inversa ma non appare evidente ad uno sguardo poco attento.

Motivo per il quale sono stati costruiti i due grafici a barre a destra. Qui appare evidente come paesi in via di sviluppo abbiano dei valori di fertility rate molto più elevati (addirittura più di 3 in Africa). Tra i paesi che spiccano troviamo soprattutto paesi africani, dell'America centrale/latina e dell'Asia.

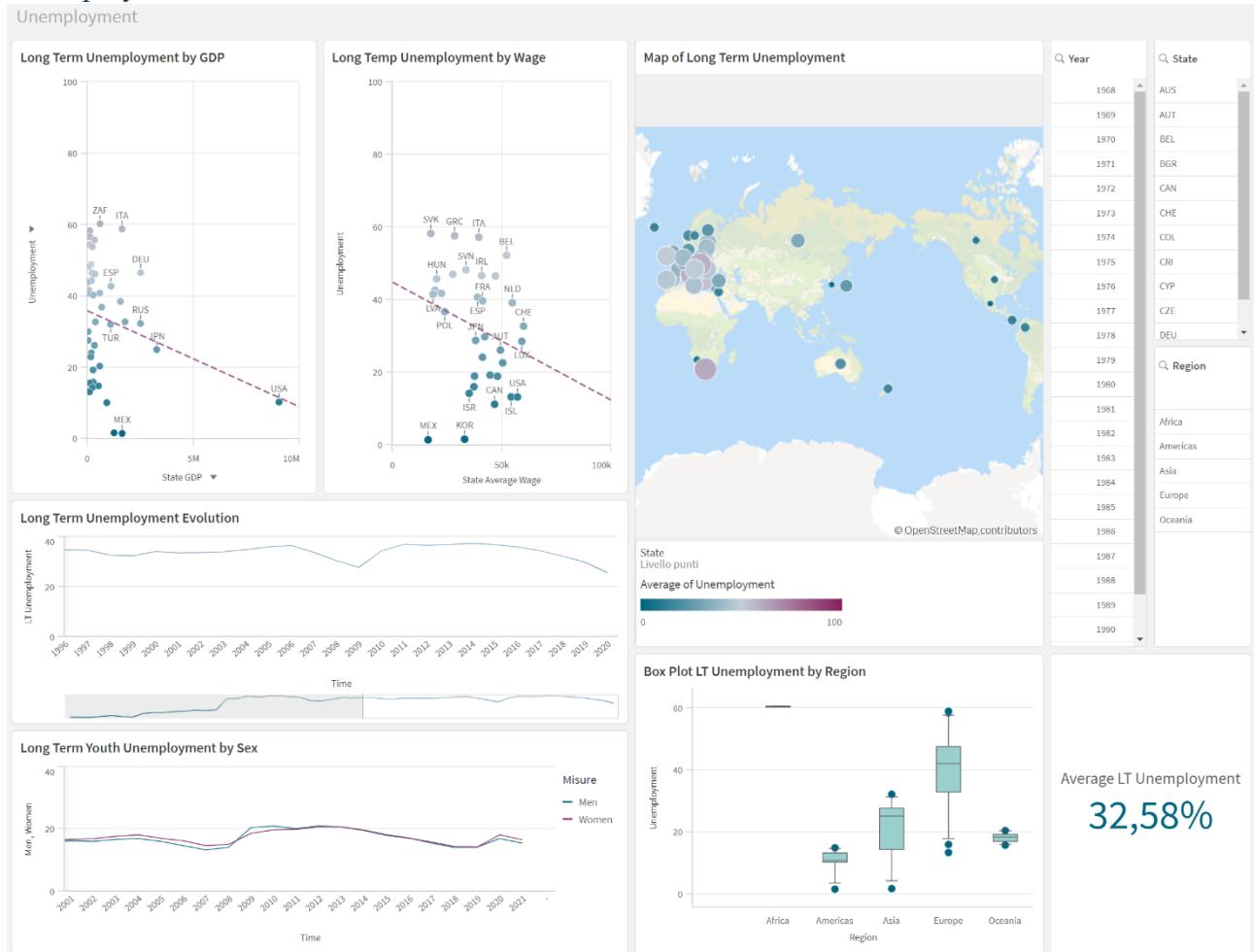
Il grafico centrale ci mostra un andamento decrescente sulle nascite. Anche in questo caso l'assenza di molti paesi africani ha il suo impatto sul valore medio ma il trend risulta comunque piuttosto chiaro.

La KPI in basso a sinistra conferma quanto detto e se viene selezionata l'Europa si ottiene un valore di 1.69 figli per madre, molto distante dalla media storica globale di 2.37. L'Italia è in linea con i valori europei.

Questo allineamento tra Italia ed Europa è giustificato dal boxplot in quanto c'è poca variabilità all'interno dei dati del vecchio continente. In Asia, invece, i valori sono spalmati ina finestra molto più ampia e paesi come l'India raggiungono valori di 4.10 figli per madre. Sappiamo bene come negli ultimi decenni questo paese ha visto una crescita demografica importante che l'ha portato ad essere, assieme alla Cina, il paese più popolato al mondo.

Infine, la mappa conferma la tendenza dell'Europa ad avere valori di fertility rate molto bassi.

Unemployment



In questo foglio sono stati analizzati trend, correlazioni ed altri aspetti riguardanti la disoccupazione.

Innanzitutto, bisogna comprendere cosa significa Long Term Unemployment e prendiamo la definizione che è presente sul sito dell' OECD: *Long-term unemployment refers to people who have been unemployed for 12 months or more.*

Vediamo dai primi due scatterplot che vi è una correlazione inversa, anche se non è di immediata comprensione in assenza di linee di trend, tra la disoccupazione di lungo periodo ed il GDP//GDP procapite/salario medio.

Dalla mappa e dal boxplot in basso a destra si può dedurre che in Europa c'è un tasso di disoccupazione mediamente più alto rispetto ad altre aree del mondo (l'Africa non viene valutata a causa dei pochi valori a disposizione).

Nei grafici a linee che troviamo sulla sinistra del foglio appare evidente come ci sia stato un aumento della disoccupazione nei primi anni 80 e che non ci siano sostanziali differenze tra la disoccupazione giovanile femminile e maschile. Solo negli ultimi anni si può osservare una leggera divergenza a sfavore delle donne.

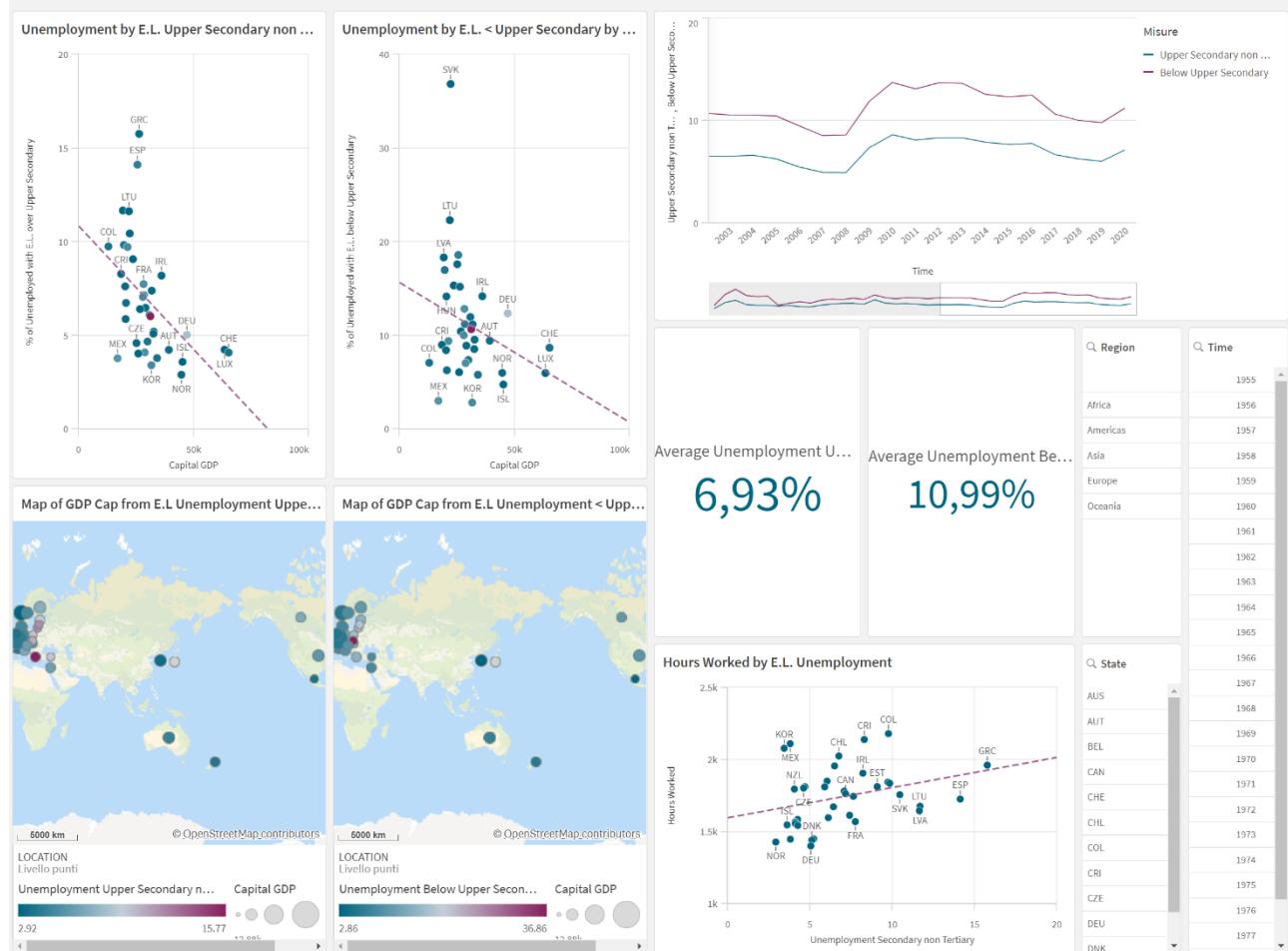
Infine, nel grafico KPI viene riportato il valore della disoccupazione di lungo periodo. Se selezioniamo Italia ancora una volta vediamo una situazione peggiore rispetto alla media globale. Il colore del valore KPI diventa, dunque, rosso, perché è stata superata una soglia critica. La figura seguente ne è una dimostrazione.

Average LT Unemployment

58,84%

Education

Education Level



In questo quarto foglio si tratta la tematica dell'educazione.

Nei primi due scatterplot appare chiara una relazione inversa tra la disoccupazione ed il GDP. I valori del primo grafico, che corrisponde alla disoccupazione delle persone con un grado di istruzione superiore alla scuola secondaria superiore, appagliono decisamente più bassi rispetto al secondo grafico, che corrisponde alla disoccupazione delle persone con un grado di istruzione inferiore alla scuola secondaria superiore.

La conferma si ha anche nel grafico a linee a destra in cui non solo questa ipotesi viene verificata per ogni anno ma la differenza tra i valori è molto significativa.

Le KPI mostrano valori medi di disoccupazione di 6.93% nel primo caso e di 10.99% nel secondo caso. In Italia questa differenza risulta comunque marcata ma si assottiglia leggermente. Si può dedurre che il grado di istruzione viene valorizzato di meno in Italia rispetto a quanto avviene in altri paesi.

Di particolare interesse è l'ultimo grafico perché viene messo in luce come nei paesi in cui si lavora per meno ore vi sia un tasso di disoccupazione delle persone con un livello di istruzione secondaria nettamente più basso.

Italy



L'ultimo foglio vuole porre l'accento su alcuni trend riguardanti l'Italia.

Vi è una crescita del GDP costante nel tempo con una lieve flessione dovuta alla crisi del 2008 e al covid nel 2020. È necessario però un appunto, cioè che i valori di GDP che abbiamo a disposizione sono superiori ai valori che si trovano in altri dataset. Inoltre, nei grafici che appaiono in seguito ad una breve ricerca su Google è presente una flessione del GDP da qualche anno, elemento che non è visibile nel grafico appena citato. Purtroppo, questo dipende dal dataset a nostra disposizione.

La fertilità è in calo negli ultimi anni ed è in linea con gli allarmi lanciati da alcuni istituti di ricerca sull'invecchiamento della popolazione italiana e sul calo demografico.

Dal grafico in basso a sinistra è visibile un aumento importante della disoccupazione a seguito della crisi del 2008.

Nell'ultimo grafico è visibile un rallentamento della crescita degli stipendi negli ultimi anni. Da escludere il valore del 2020 che potrebbe essere influenzato e distorto dalla situazione legata alla pandemia. Un aumento dei salari è avvenuto però negli anni successivi all'adozione dell'euro come moneta unica nel 2002.

Tableau Desktop

Introduzione allo strumento

Fondata nel 2003 con il nome di Tableau Software, Tableau ha sempre avuto come missione quella di rendere i dati più accessibili a tutti.

Ha concretizzato questo sogno progettando e migliorando anno dopo anno una piattaforma di Data Visualization che consente di comprendere i dati in forma visuale attraverso un'interfaccia semplice da usare e dashboard di dati interattive e intuitive.

In questo ultimo aspetto differisce rispetto a Qlik in quanto ci si focalizza di volta in volta su un foglio contenente un solo grafico. Questo permette di ottenere visualizzazioni di qualità e complessità mediamente superiori. Solo nel momento della creazione di dashboard si uniscono i grafici creati in base all'appartenenza a categorie simili.

Tableau è uno strumento molto potente che permette di svolgere anche operazioni di predizione.

Per il progetto è stata utilizzata la versione Desktop 2022.

Inserimento dataset e preparazione

L'inserimento dei file in Tableau risulta semplice, ma per creare delle relazioni è stata necessaria una fase di ETL.

Innanzitutto, per ciascuna tabella, sono state eliminate le colonne meno interessanti e modificato il tipo ad altre. In particolare, dove presente il campo location è stato settato il tipo come stringa ma con ruolo geografico CBSA/MSA (Stati Uniti), dove presente il campo Time è stato settato il tipo a Data ed infine, dove presente il campo Value è stato settato il tipo a Numero (decimale) ed è stato impostato come misura.

A causa della formattazione di questo ultimo capo nelle impostazioni locali di ciascun file è stato impostata come lingua l'inglese (Stati Uniti). In questo modo tutti i valori relativi a Value vengono letti ed interpretati correttamente.

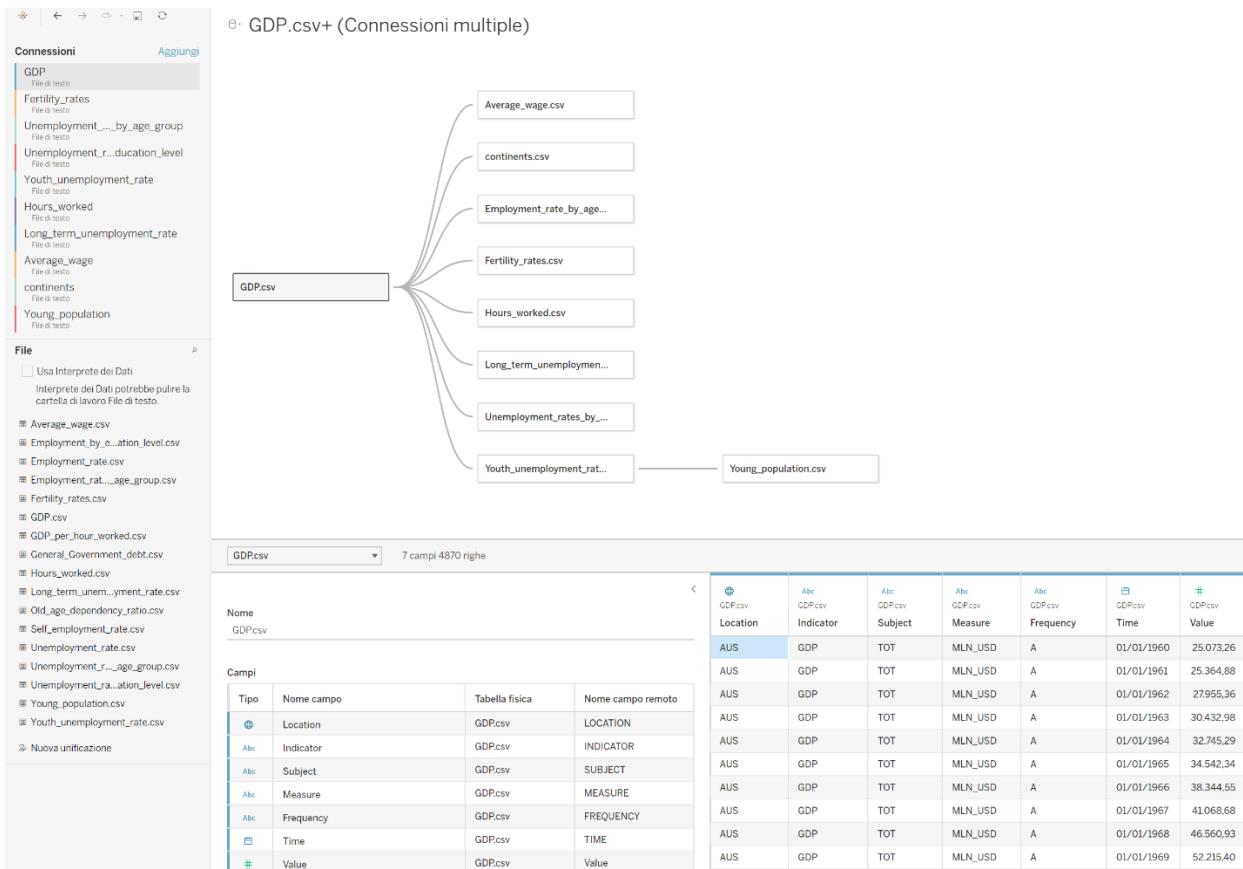
Per la creazione delle relazioni si è deciso di concatenare il campo location con il campo Time di ciascuna tabella. Per poter operare una correlazione è stato necessario eseguire un parsing del campo Time in Stringa. Un esempio di relazione tra GDP e Average_wage:

The screenshot shows the Tableau relationship editor interface. At the top, there is a dropdown menu labeled "GDP.csv — Average_wage.csv". Below it, a question asks "Qual è la differenza tra le relazioni e i join? Ulteriori informazioni". The main area displays a join configuration between two tables:

GDP.csv	Operatore	Average_wage.csv
[LOCATION]+STR([TIME])	=	[LOCATION (Average_wage.csv)]+Str([TIME (Average_wage.csv)])

Anche in questo caso, come visto per Qlik, è stata aggiunta la tabella continents.

Il risultato della fase di ETL e di creazione delle relazioni si può osservare nella seguente figura:



La connessione è stata settata *Live* e non sono state separate le tabelle come nel caso di Qlik in quanto si vuole sfruttare il filtraggio dei dati in Tableau, decisamente potente e funzionale.

Costruzione fogli di lavoro

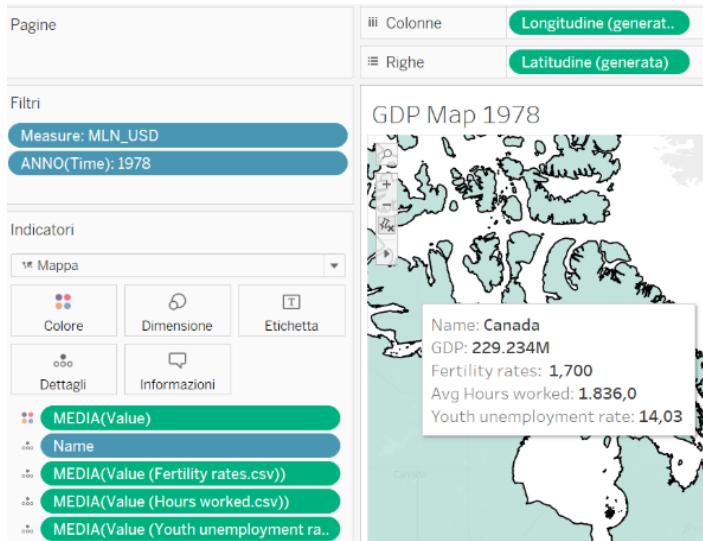
L'approccio di Tableau alla costruzione dei fogli di lavoro è piuttosto differente rispetto a Qlik in quanto l'utente viene portato a costruire un solo grafico per foglio ponendo l'accento sulla qualità del grafico stesso.

Il potenziale di suddetto programma è dato dalla modalità in cui si possono creare i fogli.

Si ha a disposizione la sezione *colonne* e *righe* in cui si possono trascinare le misure e le dimensioni di interesse. In base alla tipologia ed al numero di campi che vengono importati Tableau suggerirà delle possibili rappresentazioni. Si può dunque selezionare la visualizzazione di maggior interesse dalla sezione *Mostra* in alto a destra. Altro punto di forza è la possibilità di implementare colore, dimensione, etichette, dettagli e informazioni trascinando, anche in questo caso, il campo di interesse su uno di questi.

Questa breve introduzione serve a comprendere l'iter operativo sulla costruzione dei fogli in quanto non è possibile descrivere ciascun grafico in ogni aspetto relativo alla sua creazione. Questo ovviamente per non appesantire troppo la descrizione di questa parte del progetto.

La cura che si è avuta per ciascun foglio passa anche dalla possibilità di visualizzare numerose informazioni aggiuntive passando sopra alle varie parti dei grafici con il cursore. Questa feature è apprezzabile eseguendo il progetto sul proprio pc su Tableau.



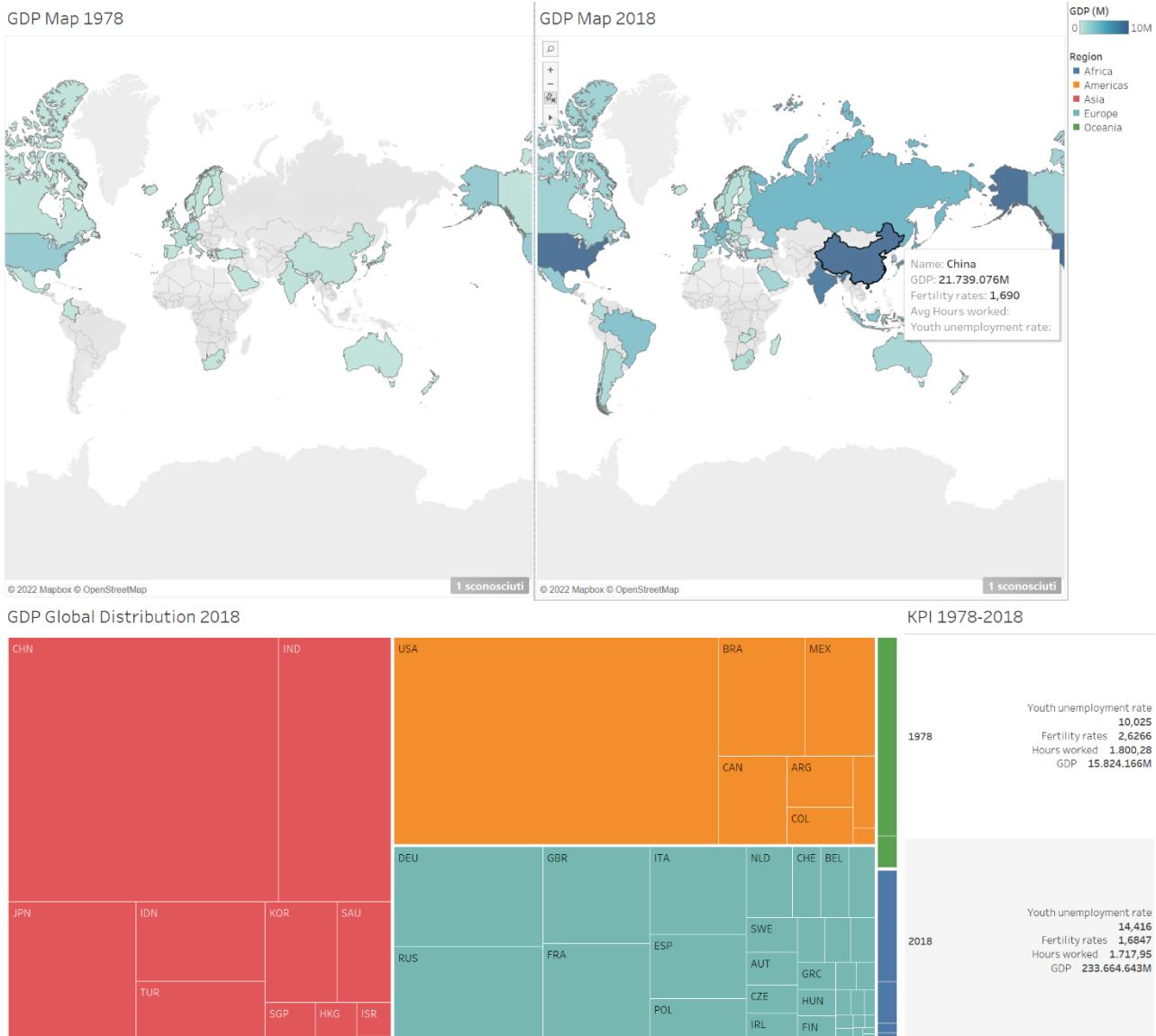
Questo piccolo esempio permette anche di giustificare la scelta di non suddividere GDP, e altri campi, in due tabelle in quanto il filtro seleziona unicamente le righe che presentano nel campo Measure il valore MLN_USD.

Allestimento dashboard

Le dashboard in Tableau sono degli aggregatori di grafici che si possono utilizzare per trattare in una unica schermata un macroargomento alla volta. Nel particolare sono state create 4 dashboard:

- Dashboard World
- Dashboard Fertility
- Dashboard Unemployment
- Dashboard Italy

Dashboard World



La prima dashboard che andiamo ad analizzare vuole confrontare alcune metriche relative ai vari stati e ai vari continenti, ponendo l'accento anche sull'evoluzione che suddette metriche hanno avuto nel tempo.

Sono stati scelti gli anni 1978 e 2018 in quanto rappresentano un periodo piuttosto lungo ma allo stesso tempo presentano un numero di dati sufficiente per compiere analisi di questo tipo. Infatti, tornando troppo indietro nel tempo molti dati non sarebbero stati valorizzati.

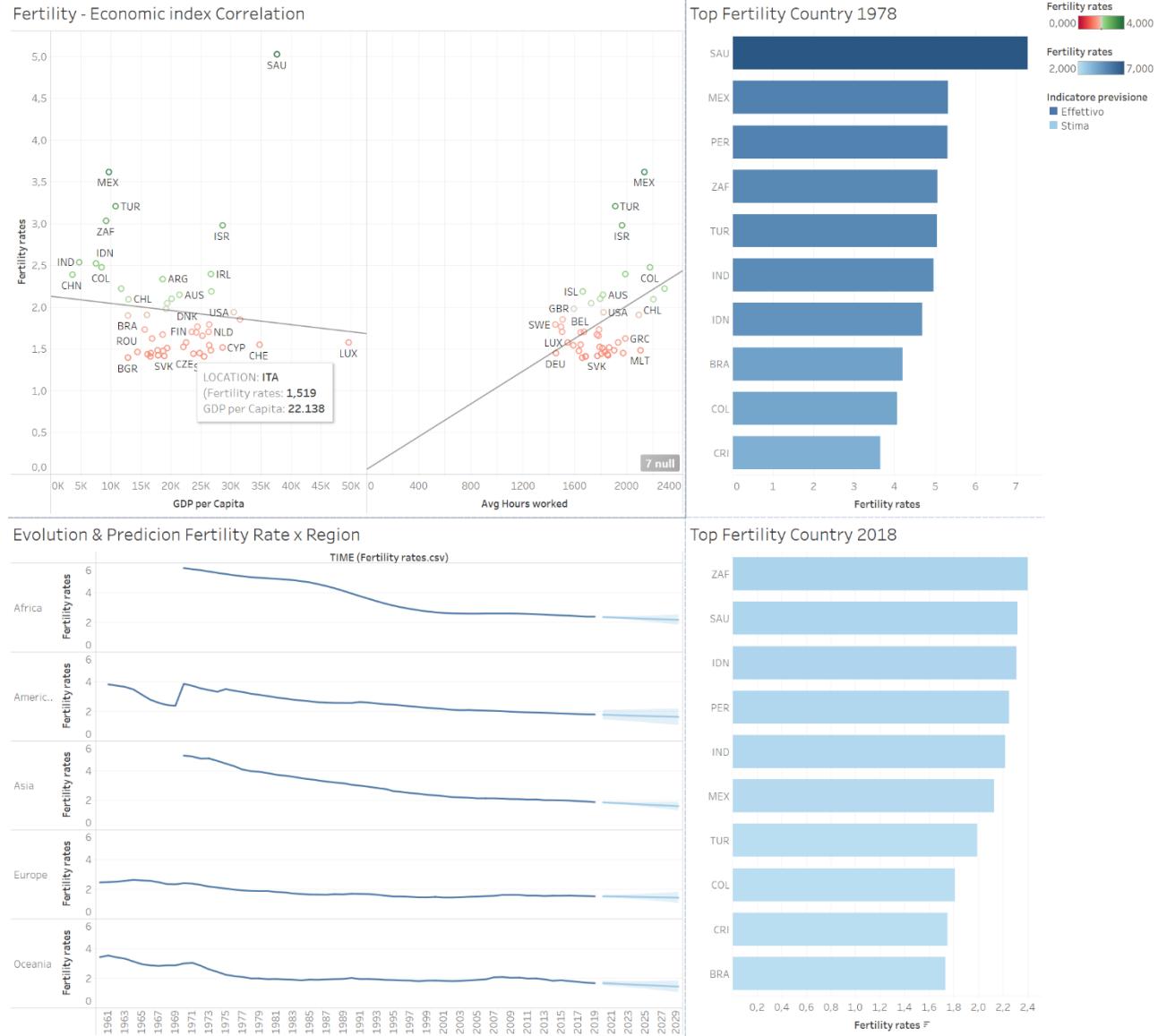
In alto possiamo osservare due mappe del globo che raccontano come in 40 anni il GDP dei vari stati è aumentato a dismisura. È evidente come nel 1978 il più grande polo economico sia rappresentato dagli USA mentre nel 2018 ad esso si affiancano paesi come la Cina e l'India. Passando sopra ogni singolo stato si possono osservare informazioni aggiuntive come il fertility rate, le ore medie di lavoro e la disoccupazione giovanile (in alcuni stati come la Cina dati di questo tipo non sono presenti).

Nel grafico ad albero invece si osserva come l'Asia nel 2018 sia il continente con il maggior peso economico, seguita da America ed Europa. L'Africa nonostante presenti un'estensione territoriale importante ha un peso veramente basso, dovuto soprattutto ad una economia poco sviluppata e basata su tecnologie arretrate. Rispetto a quanto avveniva in Qlik, qui l'interpretazione dei vari stati viene eseguita nel modo corretto senza il bisogno di preprocessing aggiuntivo, di conseguenza sono presenti la maggior parte degli stati del mondo e si ottiene un quadro molto più attinente alla realtà.

La KPI in basso a destra ci fa comprendere come a livello mondiale sia aumentata la disoccupazione giovanile, si sia abbassato drasticamente il fertility rate, si lavora mediamente meno ore ma allo stesso tempo il valore di ciò che si produce è immensamente maggiore, si passa da un GDP di 15.824.166 milioni di dollari ad un GDP di 233.664.643 milioni di dollari.

Dashboard Fertility

Fertility - Economic index Correlation



Questa dashboard ha l'obiettivo di mettere in relazione il fertility rate con altre variabili e di vedere come si evolve questo parametro nelle diverse aree del mondo.

Sugli scatterplot non ci spendiamo molte parole in quanto si ha semplicemente una conferma di quanto ricavato in Qlik.

I due grafici a barre a destra permettono di confrontare i top 10 strati per fertility rate nel 1978 e nel 2018. La differenza sostanziale di questi valori è evidente se si pensa che si è passati da più di 7 figli per madre a 2.4 figli per madre per il primo posto in classifica. La legenda ed il colore rendono evidente quanto detto.

Fino ad ora abbiamo eseguito analisi prevalentemente descrittive e diagnostiche, ma con il grafico in basso a sinistra ci si può spingere oltre e provare a predire l'evoluzione del fertility rate nel tempo in ciascun continente. Alcune metriche relative alla qualità della previsione non sono eccelse ma risulta comunque chiaro come nei prossimi anni si assisterà con molta probabilità ad un ulteriore calo delle nascite. L'Asia e

l'Africa sono storicamente i continenti con una crescita demografica maggiore ma si può notare come negli ultimi decenni questo fenomeno si sia notevolmente ridimensionato. Nonostante ciò, l'Africa potrebbe essere l'unico continente con un valore al di sopra del 2 nel 2029. L'Europa non subisce fluttuazioni importanti nel tempo ma rimane abbondantemente sotto il valore soglia, dunque, ci continuerà ad essere una decrescita nella popolazione. Spesso si associa il calo demografico ad una qualità della vita migliore e ad un paese più sviluppato dal punto di vista economico e tecnologico, queste potrebbero essere alcune delle cause che portano una diminuzione della fertility rate a livello globale, in particolar modo potrebbe essere vero per alcune aree dell'Africa.

L'intervallo di confidenza è settato al 95% ed alcune informazioni sono riportate nelle seguenti immagini.

Opzioni utilizzate per creare previsioni

Serie temporali: Anno di TIME (Fertility rates.csv)

Misure: Med. Value (Fertility rates.csv)

Previsione avanti: 10 anni (2020 – 2029)

Previsioni basate su: 1970 – 2019

Ignora ultimo: 1 anno (2020)

Modello stagionale: Nessuna (la ricerca di un modello stagionale nei dati annuali non è supportata)

Med. Value (Fertility rates.csv)

Riga Region	Iniziale 2020	Cambia dall'iniziale 2020 – 2029	Effetto stagione			Contributo		Qualità
			Lunghezza	Elevato	Basso	Trend	Stagione	
Africa	2,365 ± 0,064	-0,186	Nessuno			100,0%	0,0%	Ok
Americas	1,793 ± 0,307	-0,147	Nessuno			100,0%	0,0%	Scarso
Asia	1,869 ± 0,058	-0,255	Nessuno			100,0%	0,0%	Ok
Europe	1,534 ± 0,085	-0,086	Nessuno			100,0%	0,0%	Scarso
Oceania	1,685 ± 0,144	-0,222	Nessuno			100,0%	0,0%	Scarso

Med. Value (Fertility rates.csv)

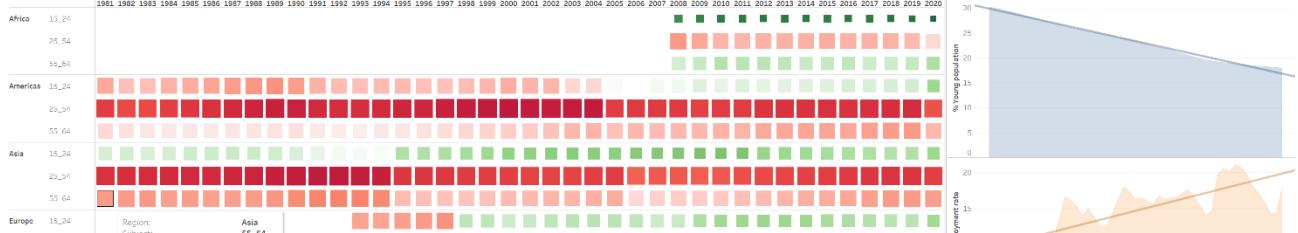
Riga Region	Livello	Modello	Metriche della qualità				Coeffienti di livellamento					
			Trend	Stagione	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Africa	Moltiplicativo	Moltiplicativo	Nessuno		0,055	0,043	0,64	1,2%	-280	0,500	0,500	0,000
Americas	Moltiplicativo	Moltiplicativo	Nessuno		0,238	0,108	1,14	3,5%	-160	0,500	0,000	0,000
Asia	Moltiplicativo	Moltiplicativo	Nessuno		0,064	0,042	0,64	1,2%	-266	0,500	0,500	0,000
Europe	Moltiplicativo	Moltiplicativo	Nessuno		0,057	0,045	1,26	2,4%	-334	0,500	0,345	0,000
Oceania	Moltiplicativo	Moltiplicativo	Nessuno		0,108	0,085	1,42	3,7%	-257	0,500	0,150	0,000

Dashboard Unemployment

Unemployment by Education Level



Employment by Age Group



Nella prima mappa risulta chiaro come il continente con la maggior numero di disoccupati sia l'Africa, con un trend in peggioramento. Seguono Europa ed America con valori piuttosto simili tra loro nell'ultimo anno. Ad un livello di istruzione crescente corrisponde una disoccupazione molto più bassa. Questo aspetto può sembrare scontato e di poco valore ma analizzando nel dettaglio i momenti storici particolarmente difficili si nota come chi ha un livello di istruzione alto non subisca variazioni significative. Per esempio, in seguito alla crisi del 2008, in Europa, per la classe Buppsry (Below Upper Secondary) si registra un +4.04%, mentre per la classe Try (Tertiary) un +1.26%.

Nella seconda mappa termica viene visualizzata, utilizzando quadrati di colore e dimensione differente, l'occupazione nelle diverse fasce di età. Ancora una volta l'Africa presenta valori decisamente negativi se rapportati agli altri continenti. Spicca in maniera positiva l'Oceania con valori positivi per ogni fascia di età. Se si analizzano le diverse fasce di età si nota una differenza importante tra la fascia 25_54 e la fascia 15_24. Questa differenza diventa sempre più accentuata nel tempo e l'America sembra fare da apripista per questo trend. Se da un lato si può giustificare la diminuzione dell'occupazione giovanile con un impegno maggiore dei ragazzi nello studio, dall'altro crediamo che non basti per giustificare un trend così marcato.

Per indagare ulteriormente sulla situazione lavorativa dei giovani sono stati sviluppati i due grafici a destra. Le linee di tendenza mostrano un aumento piuttosto importante della disoccupazione giovanile ed una diminuzione costante della percentuale di giovani sull'intera popolazione. Quest'ultimo aspetto evidenzia un invecchiamento della popolazione mondiale che era già emerso in alcune analisi precedenti.

Alcune informazioni aggiuntive sulle linee di tendenza degli ultimi due grafici:

Grafico blu

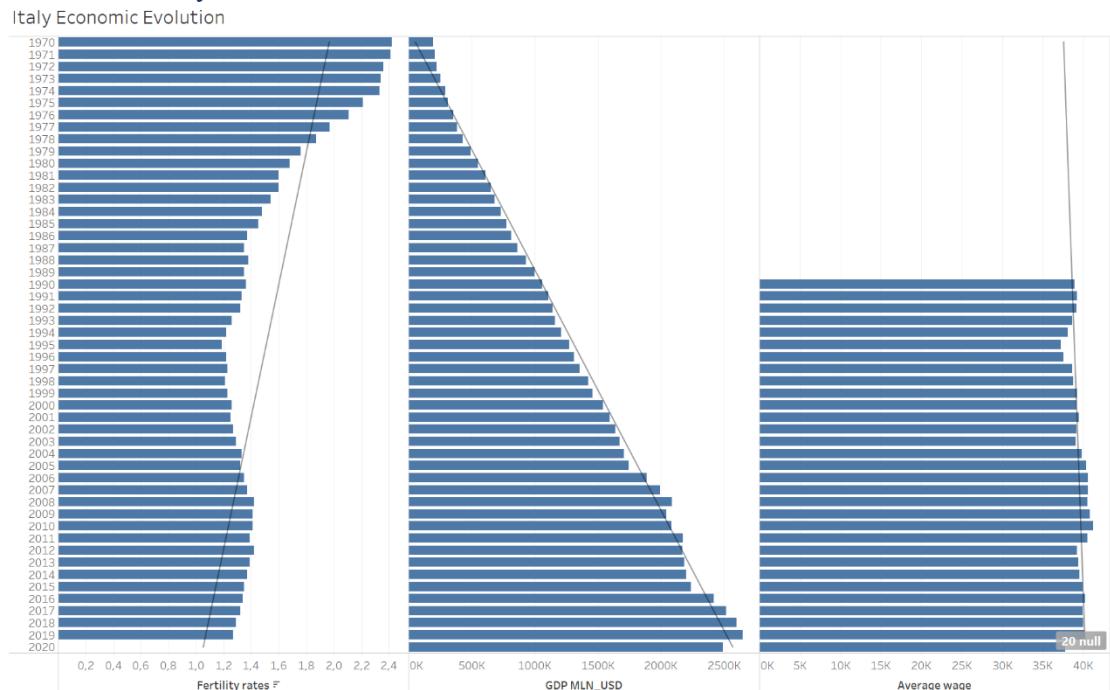
Med. Value (Young population.csv)= $-0,000713427 \cdot \text{Anno}$
di TIME (Young population.csv)+48,0835
Coefficiente di determinazione: 0,990849
Valore P: < 0,0001

Grafico arancione

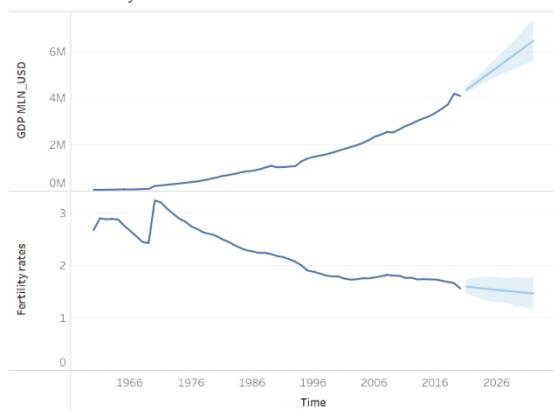
Med. Value (Youth unemployment rate.csv)= $0,000569788 \cdot \text{Anno}$
di TIME (Young population.csv)+5,1372
Coefficiente di determinazione: 0,643645
Valore P: < 0,0001

Dashboard Italy

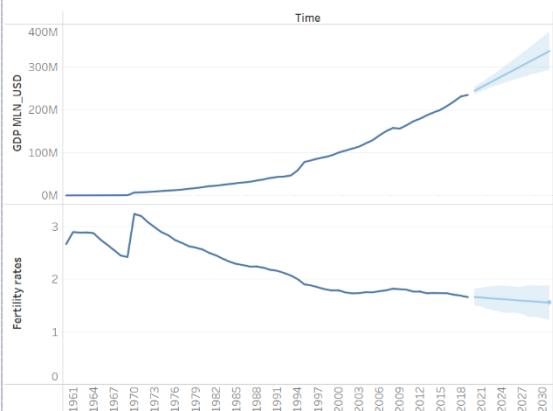
Italy Economic Evolution



Prevision Italy



Prevision World



Indicatore previsione
■ Effettivo
■ Stima

L'ultima dashboard racconta l'evoluzione di alcuni parametri importanti riguardanti l'Italia e li confronta con i corrispettivi globali.

Il primo grafico a barre mostra come il fertility rate è in calo nel corso degli anni, semmai pur con un andamento irregolare. La linea di trend essendo lineare non riesce a seguire bene l'andamento del grafico ma comunque rende l'idea di un quasi dimezzamento dei valori che si hanno oggi rispetto agli anni 70.

Il grafico a barre centrale mostra un costante aumento del GDP ma come detto in precedenza ci sono dei valori che non corrispondono a quanto trovato nel web. Se si esclude questa sovrastima di valori, che riguarda in particolar modo gli ultimi anni, si osserva una variazione negativa del GDP nell'2020, anno caratterizzato dalla pandemia di Covid-19.

Nell'ultimo grafico a barre viene indicato lo stipendio medio ed è caratterizzato da valori costanti nel tempo e sempre nell'intorno dei 39000. Proprio nell'ultimo periodo si sta discutendo della mancata crescita degli stipendi medi italiani, soprattutto se paragonati agli altri paesi dell'UE. Il tutto ha una connotazione ancora più negativa se si considera che, successivamente ai dati presenti e caricati all'interno di questo progetto, a causa del Covid e della guerra in Ucraina il prezzo di molte materie prime, e non solo, è aumentato sensibilmente.

La parte bassa della dashboard vuole confrontare l'andamento del GDP e del fertility dell'Italia con il resto del mondo. Si nota una certa correlazione per tutti e due i valori di interesse, ma come sempre è doveroso specificare che alcuni valori riguardanti l'Italia sono sovrastimati rispetto ad indici diversi da quelli dell'OECD quindi la situazione potrebbe essere mediamente peggiore.

Dal punto di vista statistico riportiamo delle informazioni interessanti sulle previsioni dedotte.

Italia

Opzioni utilizzate per creare previsioni

Serie temporali: Anno di Time
Misure: Med. Value (Fertility rates.csv),Med. Value
Previsione avanti: 12 anni (2021 – 2032)
Previsioni basate su: 1961 – 2020
Ignora ultimo: 1 anno (2021)
Modello stagione: Nessuna (la ricerca di un modello stagionale nei dati annuali non è supportata)

Med. Value (Fertility rates.csv)

Iniziale 2021	Cambia dall'iniziale 2021 – 2032	Effetto stagione Elevato Basso	Contributo Trend Stagione	Qualità
1.594 ± 0,148	-0,132	Nessuno	100,0% 0,0%	Sciarso

Med. Value

Iniziale 2021	Cambia dall'iniziale 2021 – 2032	Effetto stagione Elevato Basso	Contributo Trend Stagione	Qualità
4.354.708 ± 146.742	2.116.169	Nessuno	100,0% 0,0%	Ok

Med. Value (Fertility rates.csv)

Livello	Modello			Metriche della qualità					Coeffienti di livellamento		
	Trend	Stagione	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma	
Moltiplicativo	Moltiplicativo	Nessuno	0,129	0,067	1,24	2,7%	-236	0,500	0,000	0,000	

Med. Value

Livello	Modello			Metriche della qualità					Coeffienti di livellamento		
	Trend	Stagione	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma	
Aggiuntivo	Aggiuntivo	Nessuno	74.870	47.692	0,64	14,7%	1.357	0,500	0,395	0,000	

Mondo

Opzioni utilizzate per creare previsioni

Serie temporali: Anno di Time
Misure: Med. Value (Fertility rates.csv),Somma di Value
Previsione avanti: 12 anni (2020 – 2031)
Previsioni basate su: 1960 – 2019
Ignora ultimo: 1 anno (2020)
Modello stagione: Nessuna (la ricerca di un modello stagionale nei dati annuali non è supportata)

Med. Value (Fertility rates.csv)

Iniziale 2020	Cambia dall'iniziale 2020 – 2031	Effetto stagione Elevato Basso	Contributo Trend Stagione	Qualità
1.661 ± 0,156	-0,107	Nessuno	100,0% 0,0%	Sciarso

Somma di Value

Iniziale 2020	Cambia dall'iniziale 2020 – 2031	Effetto stagione Elevato Basso	Contributo Trend Stagione	Qualità
244.228.982 ± 8.298.243	92.546.349	Nessuno	100,0% 0,0%	Ok

Med. Value (Fertility rates.csv)

Livello	Modello			Metriche della qualità					Coeffienti di livellamento		
	Trend	Stagione	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma	
Moltiplicativo	Moltiplicativo	Nessuno	0,132	0,075	1,34	3,0%	-233	0,500	0,000	0,000	

Somma di Value

Livello	Modello			Metriche della qualità					Coeffienti di livellamento		
	Trend	Stagione	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma	
Aggiuntivo	Aggiuntivo	Nessuno	4.233.875	2.640.920	0,66	132,5%	1.841	0,500	0,329	0,000	

Power BI

Introduzione allo strumento

Power bi è una famiglia di strumenti di business intelligence di Microsoft che permette a partire dai dati di generare report e quindi informazioni a supporto delle decisioni. La condivisione delle informazioni tra gli utenti è semplificata da tutti i servizi Microsoft che gravitano attorno a questo prodotto.

L'iter operativo può essere sintetizzato in 4 punti:

- Connessione Origine Dati: Il punto di partenza per la costruzione dei report in Power Bi desktop è il collegamento all'origine dati. Definita la tipologia di fonte di dati, vengono selezionate le tabelle e caricate nel programma.
- Elaborazione Dati: Se necessario dopo aver caricato i dati, grazie all'Editor di Query è possibile operare diverse trasformazioni come rinominare campi e tabelle, eliminare ed aggiungere colonne, filtrare i dati, modificare il tipo dati, e tante altre trasformazioni. Tutte le modifiche vengono memorizzate in passaggi sequenziali che permettono di aggiungere ed eliminare delle modifiche in futuro.
- Configurazione Modello Dati: Dopo aver trasformato in modo opportuno i dati, power bi desktop mette a disposizione un ambiente per definire il modello dati. In particolare, è possibile definire le relazioni tra le diverse tabelle, creare Misure o colonne calcolate attraverso il linguaggio DAX (Data Analysis Expressions), un potente linguaggio di formule che permette di arricchire la base dati di indicatori pronti per la costruzione dei report.
- Creazione Visualizzazioni: Definito il modello dati è possibile costruire report. A partire dall'elenco dei campi e delle misure, grazie ad oggetti di visualizzazione predefiniti, possono essere costruiti i report desiderati. È disponibile anche un sezione marketplace in cui si possono scaricare ed utilizzare visualizzazioni create da Microsoft stessa, da altre società o dagli utenti.

Connessione Origine Dati

Le tabelle caricate in PowerBi sono le stesse usate in Qlik e Tableau.

Elaborazione Dati

Questo può essere considerato il punto di forza di PowerBi rispetto ai due software usati in precedenza. Infatti, quando si vuole modificare una query viene aperto il corrispettivo editor dove si possono eseguire operazioni di ETL molto avanzate.

Per comprendere meglio ciò che è stato svolto prendiamo come esempio la query `Employment_rate_by_age_group`.

La query di partenza era caratterizzata dai seguenti campi:

A _C Column1	A _C Column2	A _C Column3	A _C Column4	A _C Column5	A _C Column6	A _C Column7	A _C Column8
LOCATION	INDICATOR	SUBJECT	MEASURE	FREQUENCY	TIME	Value	Flag Codes
AUS	EMPAGE	25_54	PC_AGE	A	1979	70.71983	
AUS	EMPAGE	25_54	PC_AGE	A	1980	71.19049	
AUS	EMPAGE	25_54	PC_AGE	A	1981	71.39142	

Il riepilogo delle operazioni svolte è visibile nella sezione *Passaggi Applicati*.

Il primo passaggio definito Origine indica semplicemente il caricamento dei dati.

Anche il *filtraggio* successivo viene svolto in automatico e non ha un impatto significativo della tabella.

Con *Intestazioni alzate di livello* viene considerata la prima riga dei dati caricati come intestazione.

A causa della notazione che fa uso del punto è stato necessario trasformare il campo *Value in numero decimale ma con Impostazioni Inglese USA*.

La colonna *Flag Codes* è stata poi rimossa perché non apportava informazioni importanti al fine del progetto.

È stata creata la colonna *KEY* che sarà utile per creare relazioni tra tabelle. Tale campo è stato creato con la formula `=[LOCATION]&[TIME]`. (Es: AUS1979)

Per una visualizzazione migliore è stata ordinata la query rispetto al campo *Location* e al campo *Time*.

All'interno della colonna *Subject* erano presenti tre diverse fasce di età e ciò si ripercuoteva sul campo *KEY* che possedeva tre valori uguali. Per evitare ciò è stata eseguita una operazione di Pivot usando i nomi della colonna *Subject* e come colonna valori *Value*. Si vengono a formare dunque tre colonne con intestazione *15_24*, *25_54* e *55_64*.

Le tre operazioni successive sono operazioni di ordinamento e rinomina di campi, dunque non meritevoli di approfondimenti.

La modifica del campo data *Time in Data* necessita di una precisazione. Nell'editor di power query si può vedere come nonostante erano inizialmente presenti solo gli anni la conversione aggiunge di default anche il giorno ed il mese (1° gennaio). Per ovviare a questo problema è stato selezionato il formato appropriato ma al di fuori dell'editor, in particolare nella sezione dati di power Bi.

Infine, sono state filtrate le righe *Location* per escludere raggruppamenti di nazioni come UE, G7, ... che avrebbero portato ad errori in alcune visualizzazioni.

Il risultato finale è il seguente:

<i>#C LOCATION</i>	<i>#C INDICATOR</i>	<i>#C MEASURE</i>	<i>#C FREQUENCY</i>	<i>TIME</i>	<i>1.2 15_24</i>	<i>1.2 25_54</i>	<i>1.2 55_64</i>	<i>#C Key</i>
AUS	EMPAGE	PC_AGE	A	01/01/1979	61,32872	70,71983	43,43592	AUS1979
AUS	EMPAGE	PC_AGE	A	01/01/1980	63,1339	71,19049	43,5891	AUS1980
AUS	EMPAGE	PC_AGE	A	01/01/1981	63,91809	71,39142	42,51108	AUS1981
AUS	EMPAGE	PC_AGE	A	01/01/1982	61,81898	70,63979	40,88751	AUS1982

Per le altre query sono state eseguite operazioni simili. In particolare, operazioni di pivot riguardano:

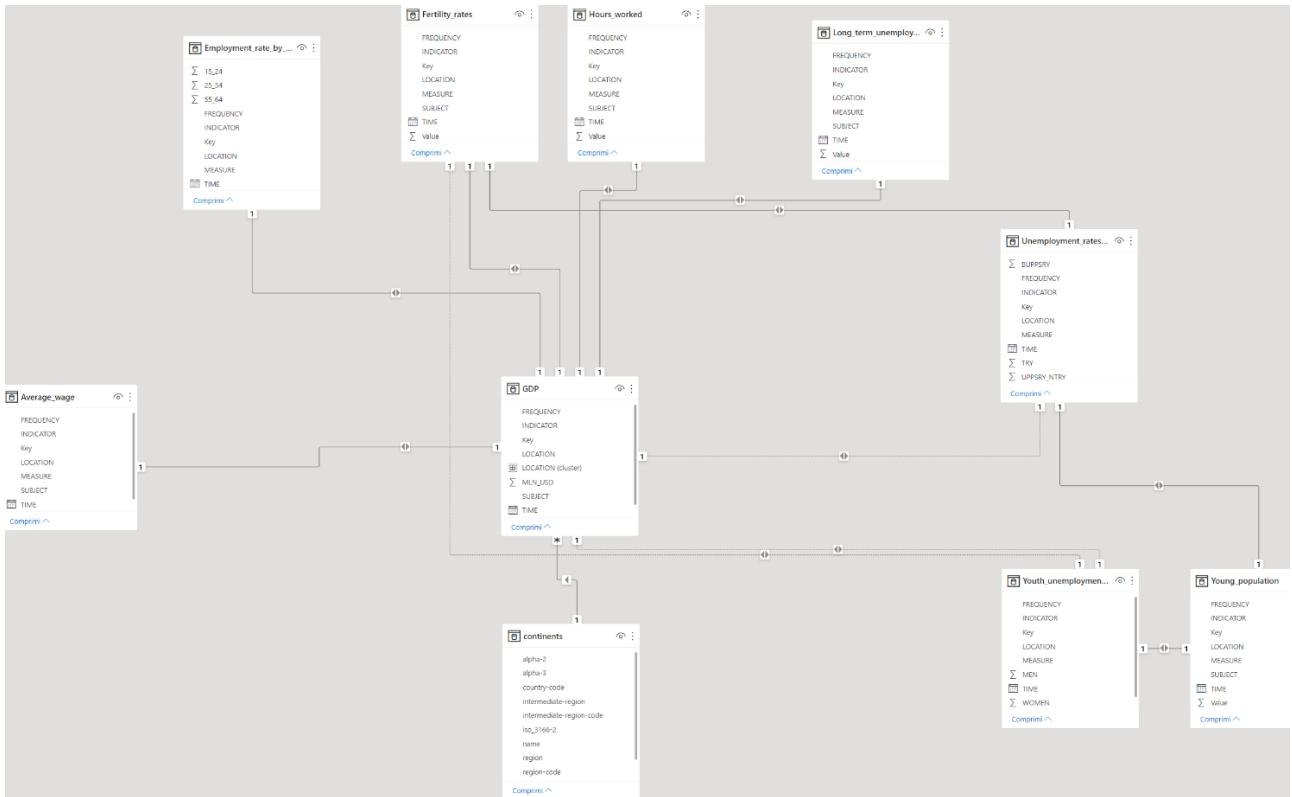
- GDP: sono state ottenute le colonne *MLN_USD* e *USD_CAP*

- Unemployment_rate_by_education_level: sono state ottenute le colonne BUPPSRY, UPPSRY_NTRY e TRY
- Youth_unemployment_rate: sono state ottenute le colonne MEN e WOMEN

Per alcune tabelle è stato necessario il filtraggio per FREQUENCY in modo da considerare solo i dati con frequenza annuale.

Configurazione Modello Dati

In questa fase sono state definite le relazioni tra tabelle. Nella quasi totalità dei casi è stato sfruttato il campo KEY costituito da LOCATION&TIME. La figura seguente mostra il risultato ottenuto.



Gestisci relazioni

Attiva	Da: tabella (colonna)	A: tabella (colonna)
<input checked="" type="checkbox"/>	GDP (Key)	Average_wage (Key)
<input checked="" type="checkbox"/>	GDP (Key)	Employment_rate_by_age_group (Key)
<input checked="" type="checkbox"/>	GDP (Key)	Fertility_rates (Key)
<input checked="" type="checkbox"/>	GDP (Key)	Hours_worked (Key)
<input checked="" type="checkbox"/>	GDP (Key)	Long_term_unemployment_rate (Key)
<input checked="" type="checkbox"/>	GDP (LOCATION)	continents (alpha-3)
<input checked="" type="checkbox"/>	Unemployment_rates_by_education_level (Key)	Fertility_rates (Key)
<input type="checkbox"/>	Unemployment_rates_by_education_level (Key)	GDP (Key)
<input checked="" type="checkbox"/>	Unemployment_rates_by_education_level (Key)	Young_population (Key)
<input type="checkbox"/>	Youth_unemployment_rate (Key)	Fertility_rates (Key)
<input type="checkbox"/>	Youth_unemployment_rate (Key)	GDP (Key)
<input checked="" type="checkbox"/>	Youth_unemployment_rate (Key)	Young_population (Key)

Nuova... Rileva automaticamente... Modifica... Elimina

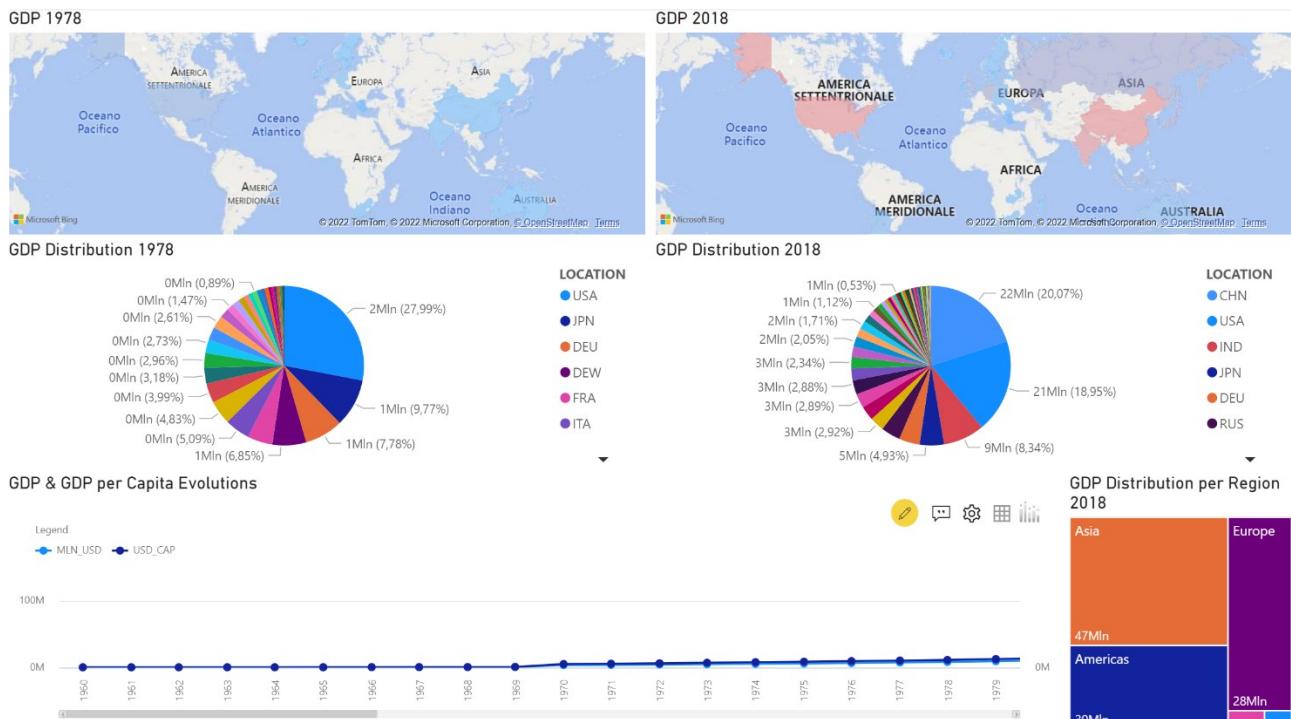
Chiudi

Creazione Visualizzazioni

In quest'ultima fase sono state create le pagine contenenti le varie visualizzazioni. In particolare, ne sono state sviluppate cinque:

- GDP
- Fertility
- Youth Population
- Un/Employment by Age Group & Instruction Level
- Italy

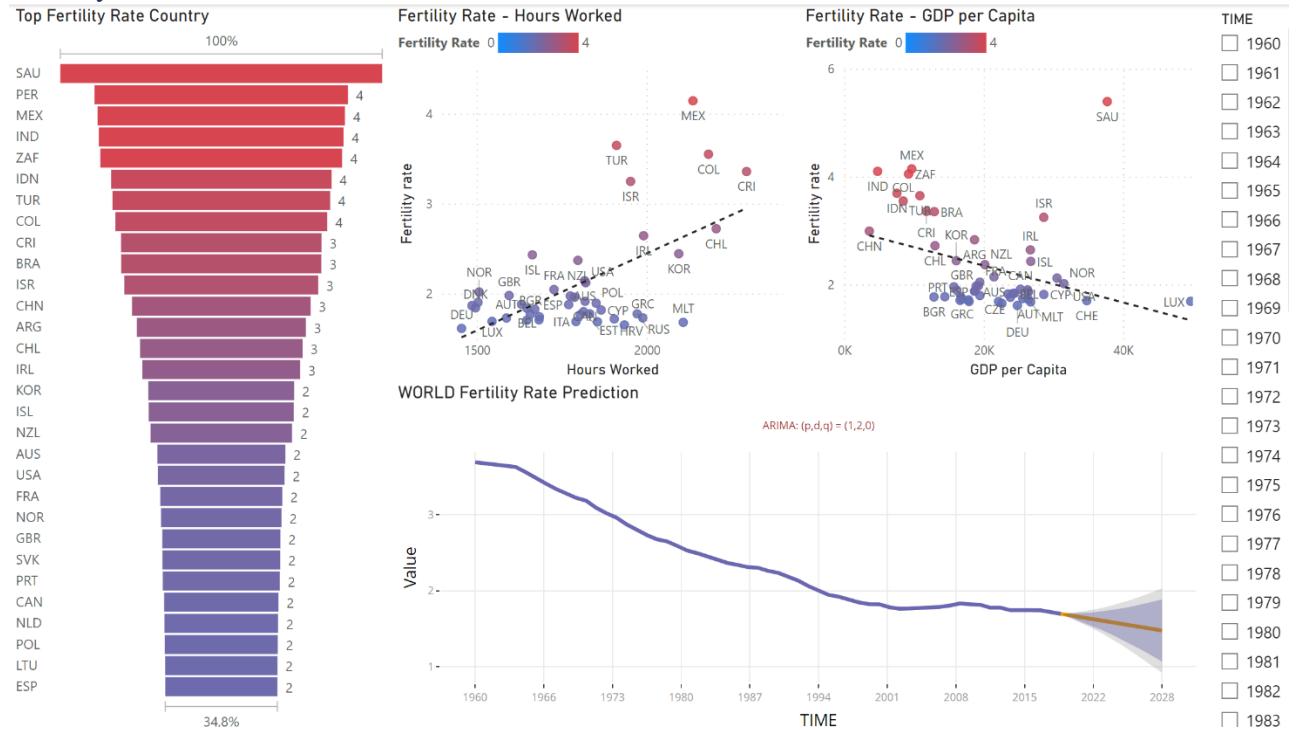
GDP



In questa prima pagina si vuole replicare quanto fatto con Tableau ovvero eseguire una comparativa su valori economici tra i vari paesi e tra il 1978 ed il 2018. I risultati ottenuti sono in linea con quanto detto discusso nella corrispettiva sezione di tableau.

Unica nota è l'utilizzo di un grafico proveniente dal marketplace ovvero l'evoluzione del GDP e del GDP pro capite. La particolarità di questo grafico è la presenza di più variabili e dunque di più assi sulle ordinate. Dalla figura non è apprezzabile l'asse alla destra del grafico a causa del ridimensionamento della finestra per l'acquisizione dell'immagine.

Fertility

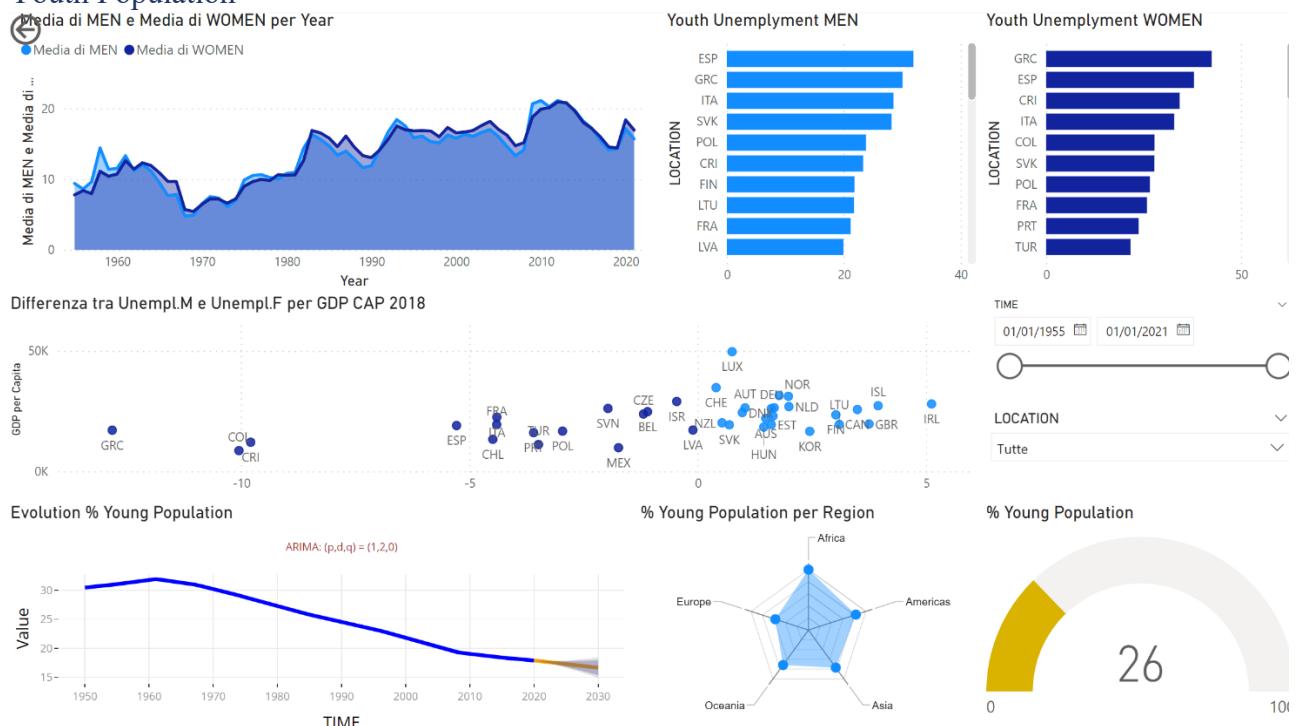


Negli scatterplot di questa pagina viene ancora una volta replicato quanto fatto in precedenza.

Nel diagramma a sinistra vengono analizzate le nazioni con il più alto fertility rate. È interessante eseguire una indagine puntuale selezionando gli anni presenti nel filtro sulla destra.

Il grafico previsionale presente in basso è stato costruito a partire dal modello ARIMA importato ancora una volta dal marketplace. Come appariva evidente anche in Tableau si prevede una ulteriore diminuzione del fertility rate nei prossimi anni. Sui valori di p,d,q non ci soffermiamo molto in quanto sarà oggetto di interesse nella relazione sulle serie temporali.

Youth Population



In questa terza pagina ricca di grafici si vogliono analizzare alcune caratteristiche relative alla popolazione giovane andando a confrontare i valori riguardanti le donne e quelli riguardanti gli uomini.

Dal primo grafico si osserva un andamento molto simile tra uomini e donne riguardo alla disoccupazione giovanile. Negli ultimi anni è presente una situazione leggermente peggiore per quest'ultime.

Analizzando il grafico a barre risulta chiaro come l'Italia sia tra i primi posti al mondo per disoccupazione giovanile per entrambi i sessi. Anche paesi come Grecia e Spagna purtroppo si ritrovano tra i primi posti di entrambi i grafici.

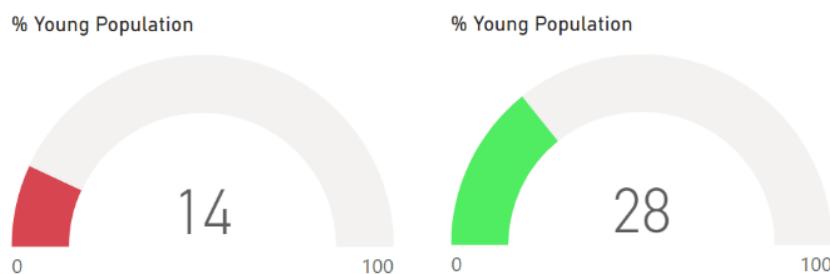
Lo scatterplot al centro è stato ottenuto creando una misura che è stata ricavata nel seguente modo: *media disoccupazione uomini – media disoccupazione donne*. L'obiettivo era verificare se in alcuni paesi vi sia una netta disparità in un senso o nell'altro. Il valore più estremo si ha in Grecia con un -12.82, dunque si deduce che in questo Paese le giovani donne hanno più difficoltà nell'inserirsi nel mondo del lavoro.

Dal grafico previsionale si deduce che dagli anni 60 ad oggi la percentuale di popolazione giovane si è quasi dimezzata e molto probabilmente il trend continuerà ad essere negativo nei prossimi anni.

Il grafico alla destra di quello previsionale è detto Radar Chart ed è stato importato dal marketplace dello strumento. È molto utile in quanto vengono confrontate le percentuali di ragazzi sull'intera popolazione nei vari continenti. L'Africa presenta una popolazione molto giovane mentre l'Europa una popolazione molto più su di età.

Infine, il misuratore permette di visualizzare in modo chiaro ed intuitivo la fetta di popolazione giovane, con un valore a livello mondiale che si assesta attorno al 26%.

Anche in questa pagina sono presenti dei filtri per data e nazione. Se selezioniamo una nazione non solo i grafici si adattano ma se vengono superati dei valori soglia cambia anche il loro colore. Vediamo l'esempio del misuratore quando si seleziona Italia (colore che diventa rosso) o Israele (colore che diventa verde).



I colori sono frutto di regole imposte nella seguente sezione:

Colore riempimento - Colori

Stile formato

Regole

Specificare il campo su cui basare questo elemento

Riepilogo

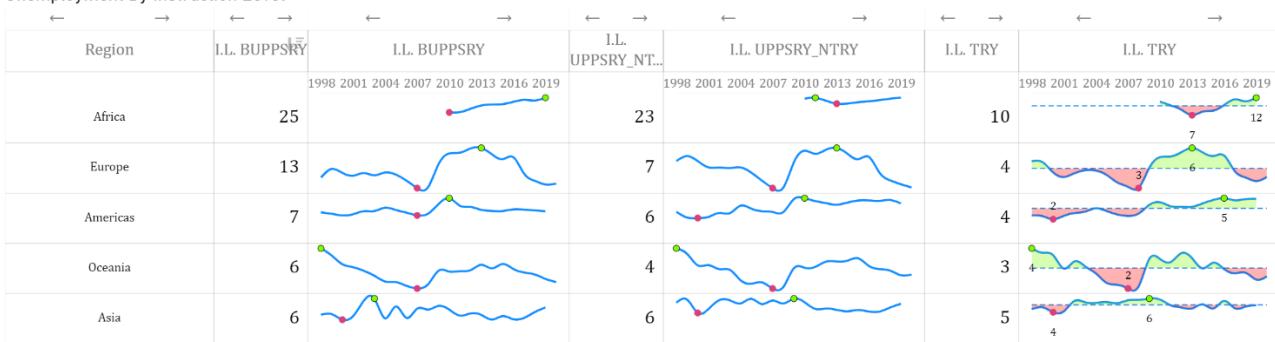
Media di Value

Media

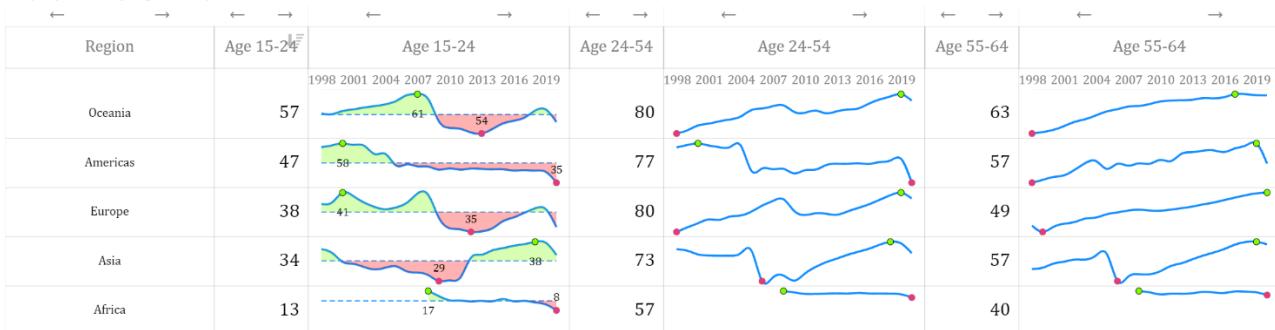
Se il valore	>=	25	Numero	e	<=	27	Numero	allora	Yellow	↑ ↓ X
Se il valore	>	27	Numero	e	<=	100	Numero	allora	Green	↑ ↓ X
Se il valore	>=	0	Numero	e	<	25	Numero	allora	Red	↑ ↓ X

Un/Employment by Age Group & instruction Level

Unemployment By Instruction Level



Employment By Age Group



Questa pagina è costituita da due grafici della stessa tipologia, *Multiple Sparklines*, importati dal marketplace.

Con tali grafici si vuole comprendere come la disoccupazione e l'occupazione siano influenzati dal livello di istruzione e dall'età del soggetto.

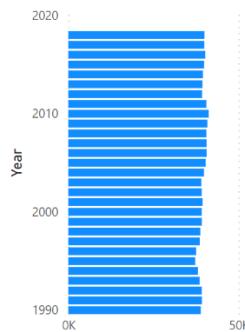
È presente una prima divisione per continente e subito sulla destra è presente un numero che caratterizza la media relativa, ad esempio, al livello di disoccupazione per le persone con un livello di istruzione del tipo BUPPSRY (Below UPPer SecondaRY). A destra di tale valore è presente un grafico con il suo andamento temporale. Il puntino verde indica il massimo, il puntino rosso il minimo. Purtroppo, a causa di un bug in tale grafico la presenza della linea che caratterizza la media ed il colore verde o rosso, a seconda del caso in cui si superi o meno tale valore, è presente in sole due colonne su sei.

Le conclusioni che si possono trarre sono che:

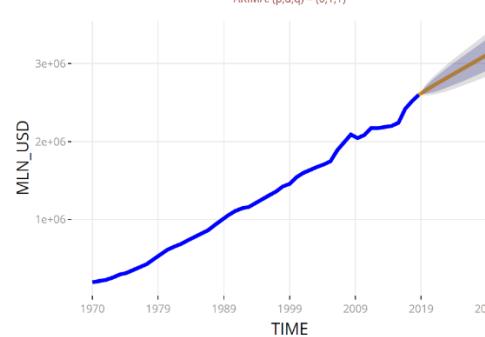
- Ad un livello di istruzione più alto corrisponde un livello di disoccupazione più basso
- L'Africa è il continente con il livello di disoccupazione più alto di tutti
- La fascia di età in cui l'occupazione è più alta è quella relativa alle persone tra i 24 e i 54 anni
- L'Oceania è il continente con il livello di occupazione più alta in due fasce su tre
- L'Africa ha il livello di occupazione peggiore per tutte le fasce di età

Italy

Average Wage ITALY



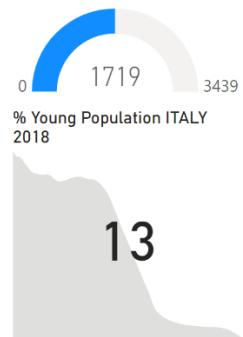
ARIMA: $(p,d,q) = (0,1,1)$



Youth Unemployment Rate Evolution ITALY (Man&Women)

2008	19	25
2009	23	29
2010	27	29
2011	27	32
2012	34	38
2013	39	41
2014	41	45
2015	39	42
2016	36	40
2017	33	37
2018	30	35

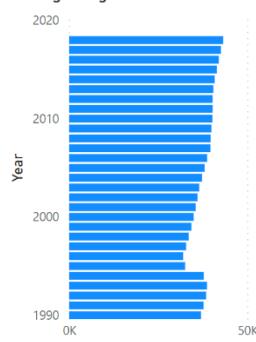
Hours Worked ITALY 2018



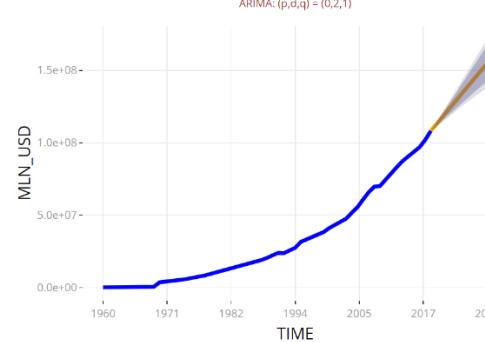
% Young Population ITALY 2018

13

Average Wage WORLD



ARIMA: $(p,d,q) = (0,2,1)$



Youth Unemployment Rate Evolution ITALY (Man&Women)

2008	14	15
2009	21	19
2010	21	20
2011	20	20
2012	21	21
2013	21	21
2014	21	21
2015	18	18
2016	17	17
2017	16	16
2018	14	15

Hours Worked WORLD 2018

18

% Young Population WORLD 2018

In questa ultima pagina si vuole ancora una volta confrontare l'Italia con il resto del mondo.

Dai grafici a barre si osserva come nel mondo si ha un aumento dello stipendio medio mentre in Italia si hanno gli stessi valori da decenni.

Dai grafici previsionali si notano aumenti sia per l'Italia che per il Mondo ma mentre nel primo caso si ha un andamento quasi lineare nel secondo si ha un andamento quasi esponenziale.

Il terzo grafico mostra come i livelli di disoccupazione siano ben più alti in Italia, sia per le donne che per gli uomini. Dato curioso è che se si guarda il dato mondiale è presente una netta differenza tra i valori pre-crisi (2008) ed il valori post-crisi, mentre in Italia c'è stato un aumento più graduale che è durato per diversi anni.

Infine, i dati relativi alle ore di lavoro, che sono paragonabili, e alla percentuale di popolazione giovane nel 2018, in cui l'Italia è decisamente sotto la media mondiale.

Considerazioni sugli strumenti utilizzati

In questa ultima sezione vorremmo fare delle considerazioni finali sui tre software utilizzati.

Qlik è risultato di facile utilizzo con una grafica piuttosto moderna e con il grande vantaggio di poter lavorare allo stesso progetto direttamente in Cloud e quindi con altre persone. Di contro un approccio online rallenta alcune operazioni come il caricamento dei dati. Una nota negativa riguarda la bassa complessità dello strumento che risulta a volte limitante per eseguire operazioni di ETL o di creazione di grafici più complessi.

Abbiamo trovato Tableau molto meccanico e schematico, che potrebbe sembrare negativo a primo impatto ma, appena si prende familiarità con lo strumento, si riescono ad eseguire operazioni di ETL di medio/alta complessità e soprattutto Tableau permette di lavorare su grafici qualitativamente superiori rispetto a Qlik e PowerBi. Siamo stati piacevolmente sorpresi dalla qualità e complessità di grafici come le mappe di calore o dalla possibilità di eseguire predizioni. In quest'ultimo caso è possibile visualizzare alcune metriche che permettono di valutare la qualità della predizione stessa.

Infine, PowerBi è risultato il software più completo a 360 gradi con una parte di etl sviluppata in modo sublime, frutto dell'esperienza di Microsoft. Operazioni come quella di pivot sono state un valore aggiunto non indifferente all'elaborazione iniziale dei dati. La fase di creazione dei grafici porta a risultati simili agli altri software ma la possibilità di sfruttare il marketplace è risultato essere un valore aggiunto da non sottovalutare. Soprattutto in quanto il marketplace consente di importare diversi strumenti che non solo consentono un'ottima analisi descrittiva, ma soprattutto consentono un primo approccio di analisi predittiva, il ché, come ben riconosciuto, aumenta esponenzialmente il valore delle analisi fatte.