

# PROGETTO DATA SCIENCE

STUDIO, ANALISI ED INTERPRETAZIONE  
MODELLO HOFSTEDE

# METODO DI LAVORO

01

## DEFINIZIONE E CREAZIONE DATASET

Creazione di un **Data Scrap** per estrarre i dati da ordinare in una tabella

02

## MODELLAZIONE DATI

Aggiunta di nuovi dati e definizione dei valori **NA**

03

## VISUALIZZAZIONE DATI

Tramite la visualizzazione dei dati si vanno a definire **relazioni tra dati** (se esistenti)

04

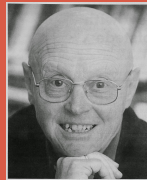
## INTERPRETAZIONI

Tramite il risultato della visualizzazione dati (e dei calcoli successivi) si va ad interpretare un **risultato**



# GEERT HOFSTEDE

Ricercatore olandese nell'ambito dello studio di **organizzazioni culturali** (pioniere nella ricerca incrociata tra gruppi culturali) di cui il suo studio più famoso che lo ha portato alla definizione del **MODELLO HOFSTEDE** che sarà la base per il progetto in questione



**“ Culture is the collective programming of the mind which distinguishes the members of one group from another ”**



# HOFSTEDE MODEL



## MASCULINITY

100

Dimensione dove **eroismo,obiettivi e raggiungimento di questi ultimi** sono fattori per la base del **successo**

0

Più il punteggio è basso e più la cultura rifiuta questi fattori sostituendoli con **modestia,cooperazione e qualità della vita**



## POWER DISTANCE

100

In questa dimensione il potere distribuito non equamente è **tollerato** (giustificazioni non richieste)

0

Il potere distribuito non equamente **non è tollerato** (giustificazioni richieste)



## INDIVIDUALISM

100

Cultura dove si dà più importanza alla mentalità dell' **io** e un focus sul **singolo** individuo

0

Si dà più importanza alla mentalità "**noi**" e focus sui gruppi che formano la cultura stessa



## LONG TERM ORIENTATION

100

**Persistenza e educazione** moderna sono fattori basilari per prepararsi al **futuro**

0

**Tradizioni e norme** (sociali e non) sono le fondamenta della cultura e qualsiasi cambiamento è visto con sospetto



## Uncertainty Avoidance

100

Cultura radicata alle **ideologie ortodosse** e non aperta a ideologie e/o movimenti non ortodossi

0

Cultura con una **preferenza per la innovazione** dove la praticità ha più valori dei principi



## INDULGENCE

100

**Norme sociali non strette** e le persone possono vivere ed esprimersi liberamente

0

**Norme sociali strette** che regolano i comportamenti dei singoli individui

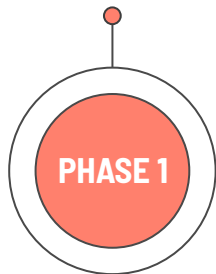
## PERCHE' IL MODELLO HOFSTEDE?

Il modello di ogni cultura viene diviso in **6 dimensioni** definite da un punteggio da 1 a 100 rendendo molto facile ed intuitivo concetti più complessi, oltretutto, permettendo di visualizzarli, e quindi interpretarli, anche attraverso i mezzi digitali dei concetti come quello della cultura

01

# DEFINIZIONE E CREAZIONE DATASET

Definizione tabella



PHASE 2

Analisi sito sorgente e  
relativo "MONSTER CODE"

Creazione Data Scrap



PHASE 3

Creazione tabella  
tramite risultato del  
Data Scrap

PHASE 4

## PHASE 1

	PD_INDEX	INV_INDEX	M_INDEX	UA_INDEX	LTO_INDEX	IND_INDEX
Country 1						
Country 2						
Country 3						
Country 4						
Country 5						
Country 6						

**Questo è il tipo di tabella che voglio ottenere quindi per ogni stato 6 indici (index) per ogni dimensione definita contenente il valore assegnato**

(la vera tabella finale sarà questa tabella unita ad un'altra presa da Kaggle tramite comando "merge" come si può vedere nel markdown)

## PHASE 2

<https://www.hofstede-insights.com/country-comparison>

HOME > COUNTRY COMPARISON

### COUNTRY COMPARISON

Select one or several countries/regions in the menu below to see the values for the 6 dimensions.

Go further, discover our cultural survey tool, the **Culture Compass™** or join our **open programme Introduction to Cross-Cultural Management**.

Type a country

Power Distance Individualism Masculinity Uncertainty Avoidance Long Term Orientation Indulgence

\* estimated

HOME > COUNTRY COMPARISON

### COUNTRY COMPARISON

Select one or several countries/regions in the menu below to see the values for the 6 dimensions.

Go further, discover our cultural survey tool, the **Culture Compass™** or join our **open programme Introduction to Cross-Cultural Management**.

Italy



\* estimated

**Da questa prima analisi notiamo che ogni stato, per accedere al suo schema contenente le informazioni che dobbiamo estrarre, ha il link uguale a quello base (come sopra) con l'aggiunta di una stringa del tipo "[nome\_stato]"**

(non tutti gli stati aggiungono semplicemente il loro nome cosa della quale ho risolto scoprendo tutti i nomi "anomali" e sono andato a sostituirli nel processo di scrapping per poi rimodificarli nella tabella finale)

<https://www.hofstede-insights.com/country-comparison/italy>

Dopo l'analisi generale sono andato a ricercare gli elementi delle specifiche pagine ottenendo il codice della lista `<ul>` contenente in ogni elemento `<li>` il nome dello stato da selezionare e successivamente ho trovato l'elemento dello schema che (purtroppo) non è un'immagine ma un canva andando a complicare un po il salvataggio automatico di ogni singolo schema (codice "data\_scrap.py" per info)

```
re-label" role="region"></div>
<div id="at4-thankyou" class="at4-thankyou at4-thankyou-background at4-hide ats-trans
parent at4-thankyou-desktop addthis-smartlayers addthis-animated fadeIn at4-show"
role="dialog" aria-labelledby="at-thankyou-label"></div>
<div id="at-expanded-menu-host"></div>
<span class="select2-container select2-container--flat select2-container--open"
style="position: absolute; top: 691.328px; left: 241.969px;">
...
<span class="select2-dropdown select2-dropdown--below" dir="ltr" style="width: 400p
x;"> == $0
<span class="select2-results">
<ul class="select2-results_options" role="listbox" aria-multiselectable="true"
id="select2-qtsw-results" aria-expanded="true" aria-hidden="false">
Go further, discov
programme Intro
span.select2-dropdown.select2-dro
down--below
400 × 378
Background
☐ #FFFFF
pass™ or join our open
ACCESSIBILITY
Name
Role
Keyboard-focusable
generic
Algeria*
Angola*
Argentina
Armenia*
Australia
Austria
Azerbaijan*
Bangladesh
Belarus*
Power Distance
Indulgence
<li class="select2-results_option" id="select2-qtsw-result-frhy-albania" role="option" aria-selected="false" data-select-id="select2-qtsw-result-frhy-albania">Albania</li>
<li class="select2-results_option" id="select2-qtsw-result-ovzi-algeria" role="option" aria-selected="false" data-select-id="select2-qtsw-result-ovzi-algeria">Algeria</li>
<li class="select2-results_option" id="select2-qtsw-result-id7r-angola" role="option" aria-selected="false" data-select-id="select2-qtsw-result-id7r-angola">Angola</li>
<li class="select2-results_option" id="select2-qtsw-result-hklm-argentina" role="option" aria-selected="false" data-select-id="select2-qtsw-result-hklm-argentina">Argentina</li>
<li class="select2-results_option" id="select2-qtsw-result-hy3d-armenia" role="option" aria-selected="false" data-select-id="select2-qtsw-result-hy3d-armenia">Armenia</li>
<li class="select2-results_option" id="select2-qtsw-result-syzw-australia" role="option" aria-selected="false" data-select-id="select2-qtsw-result-syzw-australia">Australia</li>
<li class="select2-results_option" id="select2-qtsw-result-pzym-austria" role="option" aria-selected="false" data-select-id="select2-qtsw-result-pzym-austria">Austria</li>
<li class="select2-results_option" id="select2-qtsw-result-joh7-azerbaijan" role="option" aria-selected="false" data-select-id="select2-qtsw-result-joh7-azerbaijan">Azerbaijan</li>
<li class="select2-results_option" id="select2-qtsw-result-ceky-bangladesh" role="option" aria-selected="false" data-select-id="select2-qtsw-result-ceky-bangladesh">Bangladesh</li>
<li class="select2-results_option" id="select2-qtsw-result-v2i7-belarus" role="option" aria-selected="false" data-select-id="select2-qtsw-result-v2i7-belarus">Belarus</li>
<li class="select2-results_option" id="select2-qtsw-result-otjk-belgium" role="option" aria-selected="false" data-select-id="select2-qtsw-result-otjk-belgium">Belgium</li>
<li class="select2-results_option" id="select2-qtsw-result-99r4-bhutan" role="option" aria-selected="false" data-select-id="select2-qtsw-result-99r4-bhutan">Bhutan</li>
<li class="select2-results_option" id="select2-qtsw-result-hfbc-holivia" role="option" aria-selected="false" data-select-id="select2-qtsw-result-hfbc-holivia">Holivia</li>
```

```
<canvas id="bar-chart" width="620" height="250" style="display: block;
width: 620px; height: 250px; class="chartjs-render-monitor"> == $0
```

Select one or several countries/regions in the menu below to see the values for the 6 dimensions.

Go further, discover our programmes Intro

span.select2-dropdown.select2-dropdown--below 400 × 378

Background ☐ #FFFFF

pass™ or join our open

ACCESSIBILITY

Name

Role

Keyboard-focusable

generic

Algeria\*

Angola\*

Argentina

Armenia\*

Australia

Austria

Azerbaijan\*

Bangladesh

Belarus\*

Power Distance

Indulgence





### PHASE 3

## WORK FLOW DATA SCRAP

`get_canva(x)` →

`X = country` / `return = canva of X (as data-URL)`

Tramite i driver di chrome ed un piccolo script JS il **DS** si connette alla pagina del sito contenente il canva del rispettivo stato e estrarlo come data-URL

`canva_to_image(dataurl,country)` →

Salvo con formato PNG, tramite il **dataurl**, il rispettivo canva e lo salvo con formato del tipo "[country].PNG" in una cartella predefinita

`extract_fromtxt()` →

Estraggo da un file **txt** (dove è salvato il tag `<ul>` e tutti i suoi elementi) tutti gli stati e li salvo in una lista

## PHASE 4

Molto semplicemente dopo aver ottenuto tutte le immagini per ogni singolo stato ho creato il csv (tramite file di testo poi convertito in formato .csv) copiando per ogni stato il valore delle 6 dimensioni

Ho provato a rendere parte nel processo di data scraping questa parte cercando di utilizzare **PYTESSERACT** (<https://pypi.org/project/pytesseract/>) una libreria che utilizza l'**OCR** (Optical Character Recognition) per leggere testo all'interno delle immagini ma rendeva ogni risultato o non attendibile o non standardizzato rendendo (per me) l'unica scelta quella di copiare a mano le informazioni estratte

## RISULTATO FINALE

Country <chr>	PD_INDEX <dbl>	INV_INDEX <dbl>	M_INDEX <dbl>	UA_INDEX <dbl>	LTO_INDEX <dbl>	IND_INDEX <dbl>
Albania	90	20	80	70	61	15
Algeria	80	35	35	70	26	32
Angola	83	18	20	60	15	83
Argentina	49	46	56	86	20	62
Armenia	85	22	50	88	61	25
Australia	38	90	61	51	21	71
Austria	11	55	79	70	60	63
Azerbaijan	85	22	50	88	61	22
Bangladesh	80	20	55	60	47	20
Belarus	95	25	20	95	81	15