



UNIVERSITÀ DI PISA

DEPARTMENT OF COMPUTER SCIENCE  
MASTER'S DEGREE IN DATA SCIENCE & BUSINESS INFORMATICS

---

## VA Project Report

---

Dashboard di analisi interattiva del flusso degli studenti universitari  
GitHub Project Files: [link](#).

*Author:*  
ANDREA FEDELE

3 settembre 2021

# 1 Introduzione

Il seguente report ha l'obiettivo di descrivere l'idea, i dati e la piattaforma web che è stata sviluppata per l'esame di Visual Analytics. Tale piattaforma si fonda sull'idea di voler rendere facile ed intuitiva l'analisi del modo in cui gli studenti italiani si spostano all'interno della penisola al momento della scelta della sede universitaria. Il processo che ha portato a tale dashboard parte dalla ricerca del dataset, passa da una fase di pulizia ed analisi dello stesso per terminare nello sviluppo di un applicativo web basato sul framework *Vue.js*. Le sezioni seguenti descriveranno nel dettaglio tali fasi.

## 2 Data collection e analisi dei dati

### 2.1 L'idea

L'idea di partenza sulla quale è stata sviluppata la piattaforma è quella di voler rendere interattivo e facilmente analizzabile il flusso degli studenti universitari italiani. Una dashboard sviluppata su questa idea avrebbe potuto permettere di rispondere a diverse domande di business interessanti come ad esempio:

- Qual'è la regione dalla quale vanno via il più gran numero di studenti?
- Verso dove si spostano gli studenti residenti in una determinata regione?
- Da dove arrivano gli studenti che scelgono una determinata meta?

### 2.2 Il dataset

Per rispondere a tali domande di business, ho proceduto con la ricerca del dataset facendo riferimento agli open data del Ministero dell'Istruzione e della Ricerca (MIUR) disponibili online a [questo link](#). La raccolta dei dati è stata effettuata nel settembre 2020 e, in tal data, i dati disponibili erano molto più limitanti di quelli presenti al settembre 2021. I dati presenti al momento sulla piattaforma permetterebbero, infatti, una serie di analisi e approfondimenti ulteriori che sono punto di interesse per il sottoscritto e che verranno ripresi alla fine del report nella descrizione degli sviluppi futuri. I dati raccolti, invece, non sono al momento più disponibili sul sito del MIUR perchè probabilmente integrati all'interno di altri dataset.

I dati raccolti sono stati fondamentali perchè hanno determinato le finalità e le possibilità di analisi. Il dataset raccolto era suddiviso in 3 file csv così identificati:

- Iscritti per regione dell'ateneo e provincia di residenza - a.a. 2010/11-2012/13
- Iscritti per regione dell'ateneo e provincia di residenza - a.a. 2013/14-2015/16
- Iscritti per regione dell'ateneo e provincia di residenza - a.a. 2016/17-2018/19

La totalità di questi dati permetteva di avere le seguenti informazioni relative agli anni accademici che vanno dal **2010/2011** fino al **2018/2019**: *Anno Accademico, Provincia di residenza degli studenti, Regione di residenza degli studenti, Regione dell'ateneo di destinazione, Tipologia del corso di studi, Numero di iscritti*. L'attributo che dava informazioni relativamente la provincia di residenza è stato mantenuto in un primo momento, ma è stato ritenuto poco utile nelle fasi successive perchè utilizzarlo risultava poco informativo non avendo l'informazione duale sulla provincia di destinazione.

### 2.3 Data cleaning

I 3 file che sono andati a comporre il dataset hanno subito una fase di data cleaning: non esisteva, infatti, nessuno standard relativo al naming nè delle regioni nè dei missing values. I missing values sono stati eliminati perchè impossibile recuperarli in altra maniera e perchè ritenuto scorretto ricavarli da una qualche statistica. Questa scelta è stata anche dovuta al fatto che i missing values ammontavano ad un valore irrisorio di circa 75 record per anno accademico. Inoltre, i valori stringa sono stati portati tutti in uppercase per evitare misscounting. Tutto il processo di cleaning è avvenuto off-screen e preliminarmente allo sviluppo dell'applicativo web tramite l'utilizzo del linguaggio di programmazione *Python*.

## 2.4 Data semantics

E' importante sottolineare che, nonostante l'idea iniziale fosse quella di capire verso dove si spostano gli studenti in fase di immatricolazione, il dataset contiene le informazioni relative agli *Iscritti* e non agli *Immatricolati*. Nonostante l'assenza di specifiche sul significato di *Iscritti* da parte del MIUR, è sensato pensare che comprenda non solo gli immatricolati (e cioè chi si sta iscrivendo in un determinato ateneo per la prima volta) ma anche studenti che si iscrivono agli anni accademici successivi. Ad ogni modo, tale informazione resta comunque sensata ai fini dell'obiettivo di analisi che si voleva raggiungere. Continuare ad iscriversi ad un determinato ateneo significa comunque spostarsi in tale ateneo dalla propria regione di residenza.

## 3 Web application

### 3.1 Framework e librerie di sviluppo

L'applicativo web è stato sviluppato facendo utilizzo del Framework di sviluppo **Vue.js**, che ha permesso la composizione della piattaforma in diverse componenti riutilizzabili. Relativamente al layout generale ho deciso di far utilizzo di **Buefy**, un contenitore di componenti di UI già sviluppate e pensate prettamente per Vue.js. Queste componenti sono basate a loro volta su **Bulma** che è uno fra i framework css più famosi. Tutti i chart presenti nella web application si appoggiano sulla libreria **amCharts**, ad esclusione della componente della mappa (*ChoropletMapComponent*) che è una componente custom interamente sviluppata ex novo. L'ultima libreria su cui l'applicativo si basa è **D3.js** che, oltre ad essere stata utilizzata all'interno della componente della mappa già citata per gestire il colore delle singole regioni, è stata largamente utilizzata per la lettura dei csv e l'elaborazione dei dati stessi. D3 è stata abbastanza a soddisfare tutte le elaborazioni dei dati, e non è stata necessario aggiungere ulteriori librerie per il data modelling.

### 3.2 Struttura dell'applicativo

L'applicazione è stata strutturata seguendo e sfruttando il concetto di componenti a cui il framework Vue.js tende. La web application permette due tipologie di analisi, navigabili tramite tab selection, a cui corrispondono le relative componenti Vue: analisi su singolo anno accademico (*SingleYearSearch*) e analisi su range temporale (*MultiYearSearch*). La figura 1 mostra la struttura dell'applicativo tramite la console Vue e che riflette la descrizione appena fatta.

```
▼ <Root>
  ▼ <App>
    <Header>
      ▼ <GlobalView>
        ▼ <BTabs>
          ▼ <BTabItem>
            ► <SingleYearSearch>
          ▼ <BTabItem>
            ► <MultiYearSearch>
```

Figura 1: Struttura app tramite Vue console

Per entrambe le componenti delle due Search sono stati sviluppati due file Javascript di service separati, nei quali è stata spostata completamente la logica di lettura dei dati dai csv con la conseguente elaborazione degli stessi in modo da essere conformi alle interfacce di *amCharts*. Così facendo le componenti restano totalmente agnostiche alla modalità di elaborazione, restando però fedeli ad un'interfaccia ben specifica. Inoltre, per mantenere quanto più configurabile possibile l'applicativo, la totalità dei settings è stata spostata nel file

*constants.js*: agire sulle configurazioni dell'applicativo e dei chart che lo popolano diventa semplice e poco costoso.

### 3.3 Componenti sviluppate

Le componenti Vue che compongono l'applicativo sono così suddivise:

- **4 di Layout o controllo:** Header, GlobalView, SingleYearSearch, MultiYearSearch
- **7 Componenti amCharts:** BarChartComponent, ChordDiagramComponent, HorizontalBarChartComponent, PieChartComponent, TrendLineComponent
- **2 Componenti Buefy:** TableComponent, DetailedTableView
- **1 Componente Custom:** ChoroplethMapComponent

Le componenti amCharts fanno appunto utilizzo della libreria **amCharts**, estendendo il concetto del semplice chart: tutte queste componenti, infatti, sono state sviluppate in modo tale da essere agnostiche rispetto ai dati che riceveranno e che saranno utilizzabili in diversi punti dell'applicativo senza nessun costo reimplementativo. I dati e le configurazioni vengono, infatti, trasmessi tramite le *props* di Vue come mostra la figura 2.

```
name: 'PieChartComponent',
props: {
  chartData: {
    type: Array,
    default: () => ([])
  },
  chartConfig: {
    type: Object,
    default: () => ({}),
  },
  chartId: {
    type: String
  }
},
```

Figura 2: Props su PieChartComponent

Si è fatto poi utilizzo degli altri metodi hook delle componenti offerti da Vue: l'hook *mounted* è stato utilizzato per inizializzare i chart mentre l'hook *beforeDestroy* ha permesso di distruggere le istanze Javascript generate da amCharts al momento della distruzione della componente Vue. Si è fatto inoltre utilizzo della proprietà di *watch* in modo da aggiornare i dati correttamente. Tutte queste 7 componenti sono state sviluppate seguendo in linea generale lo stesso standard così come mostrato di esempio per la PieChartComponent. Alla stessa stregua delle componenti amCharts, le 2 componenti che rappresentano delle tabelle sono state sviluppate integrando ed estendendo delle componenti Buefy. L'unica componente interamente custom che è stata sviluppata è la **ChoroplethMapComponent**. La motivazione dietro questa scelta è dovuta al fatto che le componenti *Leaflet* compatibili con Vue sono ancora o troppo limitanti o troppo elaborate. Inoltre, sviluppare interamente questa componente, ha creato il giusto spazio per poter utilizzare ed integrare *D3.js* oltre le semplici funzionalità di lettura csv.

### 3.4 Single Year Search

La ricerca per singolo anno accademico è composta da 3 sezioni: la prima riassume delle statistiche generali e non permette interazioni, la seconda permette di analizzare interattivamente e in dettaglio gli studenti **in uscita** da una (o più) determinata regione mentre l'ultima permette una analisi interattiva degli studenti **in arrivo** in una (o più) determinata regione. Tutte queste sezioni sono incluse all'interno di tab che possono

essere nascoste o mostrate a seconda della volontà dell'utente che fa uso dell'applicativo come mostra la figura 3.

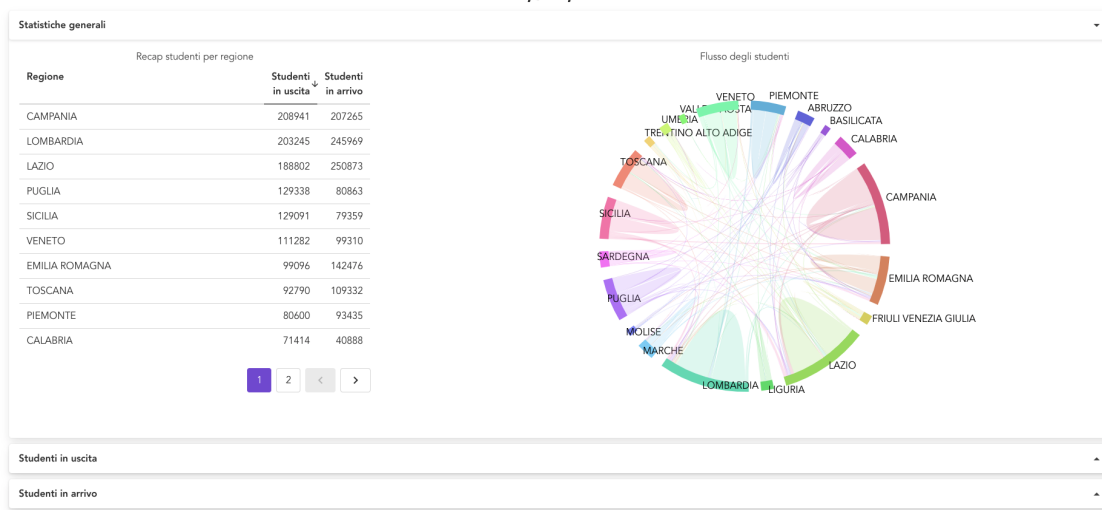


Figura 3: Single Year Search

La Single Year Search permette, come suggerisce il nome, di selezionare e filtrare l'anno accademico di interesse e di poterne analizzare i dati all'interno delle 3 sezioni appena descritte.

### 3.4.1 Statistiche generali

Questa sezione include una tabella riassuntiva che mostra, per ogni regione, il numero di studenti che vanno via e quelli che arrivano. Entrambi i campi sono ordinabili in modo crescente o decrescente a seconda delle volontà dell'utente. Oltre la tabella riassuntiva, qui abbiamo un Chord Diagram che permette di analizzare il flusso degli studenti fra regioni. E' importante sottolineare che gli studenti che risultano come uscenti da una determinata regione potrebbero anche configurarsi come entranti nella stessa: si tratta di studenti che sono rimasti all'interno della propria regione di residenza. In questa sezione tale informazione è volutamente non filtrabile: si tratta di una sezione che vuole dare una idea generale e che verrà poi approfondita nelle seguenti sezioni.

### 3.4.2 Studenti in uscita

Questa sezione permette di analizzare il flusso di studenti in uscita dalle regioni. La **ChoroplethMapComponent** è il fulcro della sezione perché, oltre a mostrare il numero di studenti uscenti sulla mappa stessa, permette di filtrare e aggiornare i dati di tutti gli altri grafi in base alla regione selezionata. Filtrando su una o più regioni tramite click su mappa, si potrà infatti andare ad aggiornare i dati del Chord Diagram, della tabella e del Bar Chart successivi. Il Chord Diagram presente in questa sezione permette di escludere dal grafo quegli archi entranti e cioè quegli archi che rappresentano gli studenti rimasti all'interno della stessa regione. La tabella presente in questa sezione permette di analizzare verso dove si sono spostati gli studenti residenti nelle regioni selezionate su mappa. Aprendo il dettaglio di ogni singola regione, infatti, sarà possibile analizzare il numero di studenti che si sono spostati in tutte le altre regioni e in che percentuale ogni regione di uscita influisce sul totale. Infine, il Bar Chart rappresenta la tipologia del titolo di studio che gli studenti attualmente filtrati sono andati a conseguire.

### 3.4.3 Studenti in arrivo

Questa sezione è identica graficamente a quella degli studenti in uscita precedentemente descritta, con la differenza che qui il focus dell'analisi sono le regioni verso le quali gli studenti si sono spostati. Filtrando

su una o più regioni questa sezione permetterà di analizzare da dove gli studenti sono arrivati nella regione selezionata.

### 3.5 Multi Year Search

La ricerca per multi anno accademico è composta da 2 sezioni: la prima riassume una serie di statistiche generali mentre la seconda permette di analizzare i trend di studenti uscenti contro quello di studenti entranti in una determinata regione. Tutti i dati delle sezioni sono controllati dallo slider presente al top della pagina e che permette di selezionare gli anni accademici di interesse all'analisi. Entrambe le sezioni sono incluse all'interno di tab che possono essere nascoste o mostrate a seconda della volontà dell'utente così come nel caso della Single Year Search.

#### 3.5.1 Statistiche generali

Questa sezione contiene un Line Chart che permette di analizzare come è cambiato il numero di studenti universitari iscritti nel corso degli anni accademici. Oltre al Line Chart, questa sezione permette di analizzare le prime 5 regioni ordinate in base al numero maggiore (o minore) di studenti in uscita. Un button switch permette di visualizzare e aggiornare i dati del Bar Chart e della Pie Chart in base al maggiore o al minor numero di studenti in uscita. La Pie Chart mostra in percentuale la destinazione verso la quale questi studenti si sono spostati suddividendoli per *stessa regione* e *altre regioni*. Una successiva ulteriore coppia di Bar Chart e Pie Chart permette invece di fare la stessa analisi però in base al numero di studenti in arrivo nelle regioni d'Italia.

#### 3.5.2 Trend studenti in uscita vs studenti in arrivo

Questa sezione presenta un Double Line Chart che permette di mettere a confronto i trend degli studenti in uscita e in arrivo in una determinata regione. Un button switch permette inoltre di escludere o includere nell'analisi quegli studenti che restano all'interno della stessa regione. Un esempio che mostra i trend a confronto per la regione Basilicata escludendo quegli studenti che restano nella stessa regione è mostrato in figura 4.

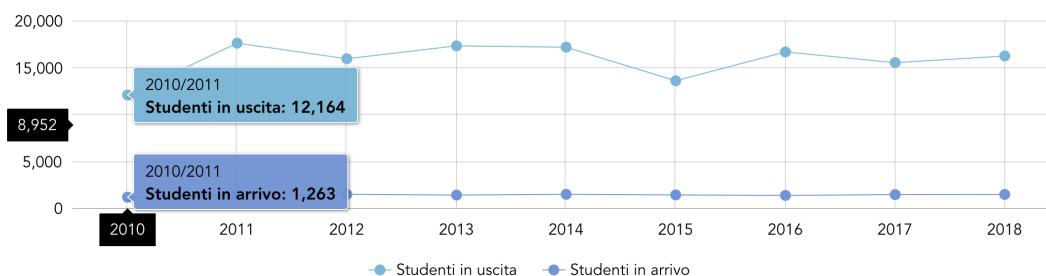


Figura 4: Trend studenti in uscita vs in entrata in Basilicata

## 4 Conclusioni e lavori futuri

La piattaforma che è stata sviluppata permette diverse analisi, che però potrebbero essere ampliate ed integrate con il corso di studi, il sesso degli studenti ed eventualmente il costo medio dell'affitto per permettere una ricerca di correlazioni tra il flusso migratorio e tali fattori. I dataset presenti ad oggi settembre 2021, e che non erano presenti al momento della data collection, permetterebbero solo in parte tali analisi e richiederebbero comunque una revisione dell'attuale piattaforma in funzione di nuove domande di business alle quali si vorrebbe cercare risposta. Da un punto di vista tecnico, oltre a poter migliorare da un punto di vista di UX/UI, la piattaforma potrebbe fare un grosso salto di qualità nel momento in cui i servizi delle due

componenti principali (Single e Multi Year Search) venissero spostati su di un server diverso da quello che serve i file statici di front end. Al momento tali servizi sono scritti in file Javascript separati e permetterebbero una facile integrazione all'interno di un server Node.js. Altro punto di interesse su cui continuare a lavorare è quello relativo ad un nuovo indice da proporre per poter fare un confronto più sensato degli studenti che si spostano. Navigando la piattaforma è spesso risultato che le regioni con un gran numero di studenti iscritti sono anche quelle con una popolazione più grande. Uno sviluppo futuro potrebbe andare a considerare un indice del tipo:

$$\frac{\text{Studenti iscritti della regione}}{\text{Popolazione tra i 18 e i 27 anni della regione nell'anno di interesse}}$$

L'autore si riserva di poter lavorare ancora su questa piattaforma per evolverla tecnicamente e analiticamente seguendo proprio questi spunti riflessivi finali.