

## PROVA SCRITTA 1

### Quesito 1:

Il candidato indichi quali metodi utilizzerebbe per gestire opportunamente serie storiche di dati relativi alla stessa grandezza, ma con indici temporali non omogenei, ad esempio alcune di esse hanno timestep di 1 minuto, altre di 15 minuti, altre di 1 h.

### Quesito 2:

Come imposterebbe l'analisi del trend di iscrizioni all'Ateneo degli studenti negli ultimi 5 anni, per individuare quali sono le categorie più presenti e quelle meno rappresentate?

## PROVA SCRITTA 2

### Quesito 1:

Avendo a disposizione un dataset con dati incompleti o non validi (missing), il candidato indichi come si comporterebbe per gestirli opportunamente in fase di analisi.

### Quesito 2:

Il candidato illustri come imposterebbe un sondaggio sugli utenti dell'Ateneo per comprendere le abitudini nell'utilizzo dei mezzi di trasporto per recarsi da casa alla sede di lavoro, al fine di individuare i fattori su cui puntare per incentivare una mobilità maggiormente sostenibile.

### PROVA SCRITTA 3

#### Quesito 1:

Avendo a disposizione un dataset con diverse serie di dati, il candidato indichi quali operazioni preliminari eseguirebbe prima di avviare le analisi.

#### Quesito 2:

Il candidato illustri come imposterebbe un questionario di customer satisfaction, per comprendere quali sono i fattori di maggiore e minore soddisfazione del personale dipendente del Politecnico di Milano.

## PROVA ORALE 1

### Quesito 1:

Il candidato proponga KPI e modelli di monitoraggio opportuni per misurare obiettivi di risparmio energetico degli edifici dell'Ateneo.

### Quesito 2:

Il candidato illustri su quali fattori baserebbe l'analisi dei trend di iscrizione di studenti provenienti dall'estero.

### Quesito 3:

Legga e traduca il seguente testo:

*Statistics involves the collection and analysis of data. Both tasks are critical. Raw data without analysis are of little value, and even a sophisticated analysis cannot extract meaningful information from data that were not collected in a sensible way.*

## PROVA ORALE 2

### Quesito 1:

Il candidato illustri quali metodi potrebbe utilizzare per individuare correlazioni e pattern interpretativi, particolari trend o modelli ricorrenti, avendo a disposizione le risposte ad un questionario relativo alle abitudini di mobilità nel trasferimento casa-lavoro con 10'000 rispondenti.

### Quesito 2:

Il candidato illustri quali indicatori prenderebbe in esame per valutare i fattori principali che portano gli studenti ad abbandonare il corso di studio nei primi 2 anni di frequenza.

### Quesito 3:

Legga e traduca il seguente testo:

*Effective data analysis requires an understanding of the research problem. We must know the goal of the research and what questions we hope to answer. It is important to have a clear direction before gathering data to lessen the chance of being unable to answer the questions of interest using the data collected.*

### PROVA ORALE 3

#### Quesito 1:

Avendo a disposizione i dati di iscrizione e di rendimento degli studenti dell'Ateneo, il candidato illustri le metodologie con cui potrebbe organizzare i dati e creare cluster per descrivere in modo sintetico la popolazione studentesca.

#### Quesito 2:

Il candidato illustri quali fattori prenderebbe in considerazione per analizzare i trend di collocamento sul lavoro degli studenti del Politecnico di Milano neolaureati e a distanza di 10 anni dalla laurea?

#### Quesito 3:

Legga e traduca il seguente testo:

*Data sets often consist of measurements collected over time at regular intervals so that we can learn about change over time. For example, stock prices, sales figures, and other socio-economic indicators might be recorded on a weekly or monthly basis.*

## PROVA ORALE 4

### Quesito 1:

Avendo a disposizione i dati di consumo energetico degli edifici dell'Ateneo, il candidato illustri quali metodologie adotterebbe per individuare edifici con consumi anomali.

### Quesito 2:

Avendo i dati di customer satisfaction relativi alla ristorazione delle mense di Ateneo, illustrare le metodologie per individuare gruppi omogenei di utenti in relazione al livello di soddisfazione.

### Quesito 3:

*When describing numerical data, it is common to report a value that is representative of the observations. Such a number describes roughly where the data are located or "centered" along the number line, and it is called a measure of center. The two most popular measures of center are the mean and the median.*