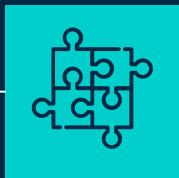


The background is a dark blue gradient. It is decorated with various geometric elements: small squares in white, teal, orange, and pink, and thin white vertical lines of varying lengths. These elements are scattered across the slide, creating a modern, data-oriented aesthetic.

# DATA SCIENCE CONSULTING

Analisi della soddisfazione dei  
clienti di una compagnia aerea

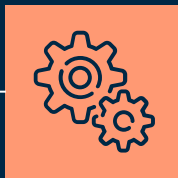
# AGENDA



01

## ANALISI

Presentazione del problema, dati a disposizione e obiettivi



02

## PROCESSO

- EDA
- Caratteristiche dei modelli progettati
- Model evaluation e feature importance



03

## CONCLUSIONI

Insights rilevanti e sviluppi futuri



# Problema e Obiettivi

La soddisfazione dei clienti è fondamentale per qualsiasi azienda, in particolare nel settore del **turismo** e dei **viaggi**.

L'**obiettivo** di questa presentazione è individuare i **fattori** che influenzano la soddisfazione dei clienti e comprendere al meglio il loro **comportamento** attraverso lo studio del campione di dati fornito.





# Descrizione del Dataset

Categoriche,  
numeriche e ordinate

**FEATURES**

**DATASET**

103.904 righe e 24  
colonne features + 1  
target



La metrica da  
predirre

**SODDISFAZIONE**

**ANALISI COMPLETA**

I dati analizzati  
riguardano anche  
il rapporto prima  
del volo



# Exploratory Data Analysis

## VIAGGIATORI OCCASIONALI

Sembra che questo tipo di viaggiatori sia mediamente contento del servizio e possiamo notarlo dalle due metriche riguardo il tipo di viaggio e la classe di prezzo scelta

### Soddisfazione

	satisfaction
Type of Travel_Personal Travel	0.449000
Class_Eco	0.451118

### Insoddisfazione o Neutralità

	satisfaction
Online boarding	0.503557
Type of Travel_Business travel	0.449000
Class_Business	0.503848

## BUSINESS

I clienti Business sembrano non soddisfatti della qualità del servizio offerto

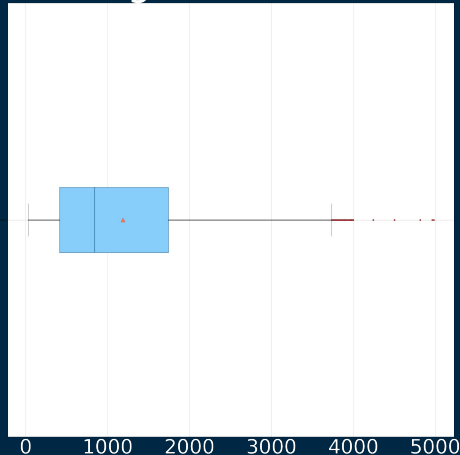
## ONLINE BOARDING

È la **seconda** correlazione più significativa del Dataset



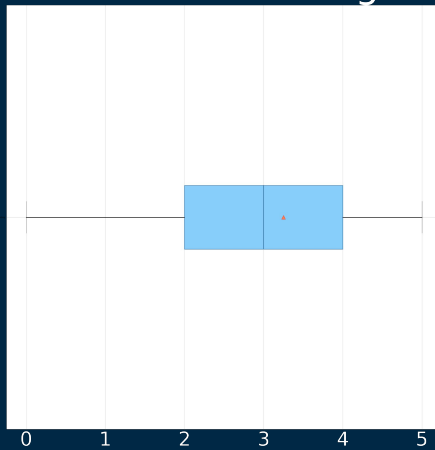
# Exploratory Data Analysis

Flight Distance



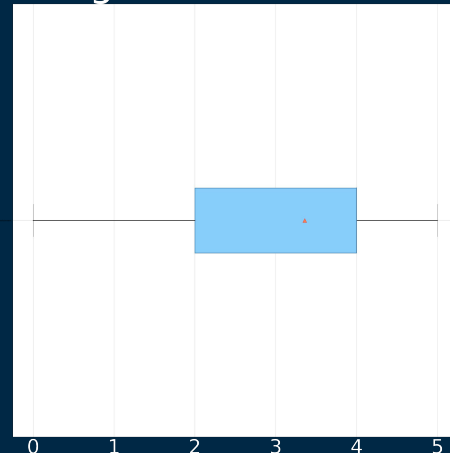
Possiamo vedere come la distanza del volo sia molto **dispersiva** arrivando a 5000, mentre la maggior parte dei dati è raggruppato sotto la metà di questo numero

Online boarding



Le metriche ordinate per voti dati mostrano un **pattern simile** dove la maggior parte dei voti ricade tra 2 e 4 e di conseguenza la media ricade vicino alla mediana, portando comunque a una **leggera asimmetria** verso un lato o l'altro

Inflight entertainment





# Modello #1

Classificazione | Ensemble Method  
(KNN + Decision Tree + Logistic Regression)

Il modello si basa sull'unione di 3 differenti algoritmi.

La model selection è avvenuta tramite **cross validation**.

Gli iper-parametri da utilizzare sono stati scelti attraverso la suddivisione del campione di train in 5 split in modo da massimizzare la **Balanced Accuracy**.

KNN

`'n_neighbors': 6,`  
`'weights': 'distance'`

DT

`'criterion': 'entropy',`  
`'max_features': None,`  
`'min_samples_leaf': 9,`  
`'splitter': 'best'`

LR

`'C': 0.1,`  
`'penalty': 'l1',`  
`'solver': 'saga'`



# Modello #2

## Classificazione | Random Forest

Il secondo modello si basa interamente su una Random Forest. Si basa sulla logica di **bagging**, gli alberi decisionali lavorano in parallelo e ognuno di loro effettua un voto di predizione. Il valore che ottiene la **maggioranza** di predizioni dai vari DT, sarà quello predetto dal Random Forest.

Similmente al modello#1 la model selection è avvenuta tramite **cross validation** e gli iper-parametri da utilizzare sono stati scelti attraverso la suddivisione del campione di train in 5 split in modo da massimizzare la **Balanced Accuracy**.

RF

```
'criterion': 'entropy',  
'min_samples_leaf': 1,  
'min_samples_split': 7
```



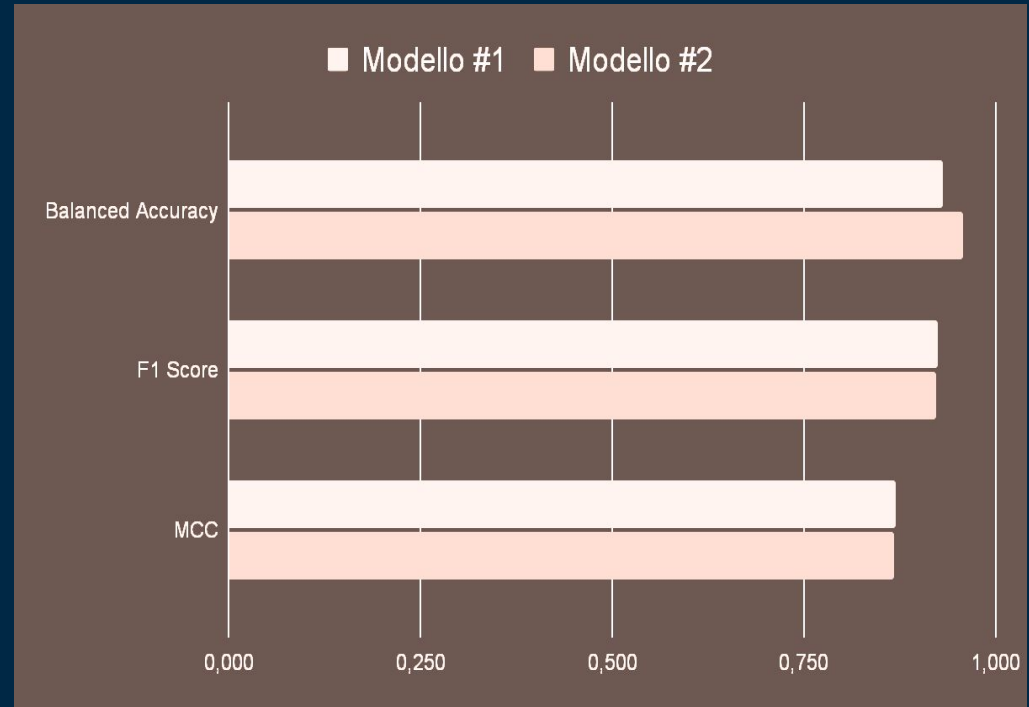


# Confronto tra Modelli

Come si può notare il modello #2 restituisce **risultati migliori** soprattutto nella Balanced Accuracy.

Una Balanced Accuracy migliore ci dice che il modello fa meno errori sia di falsi positivi che di falsi negativi.

Le altre 2 statistiche ci indicano che i 2 modelli sono simili nel risolvere il problema di classificazione.

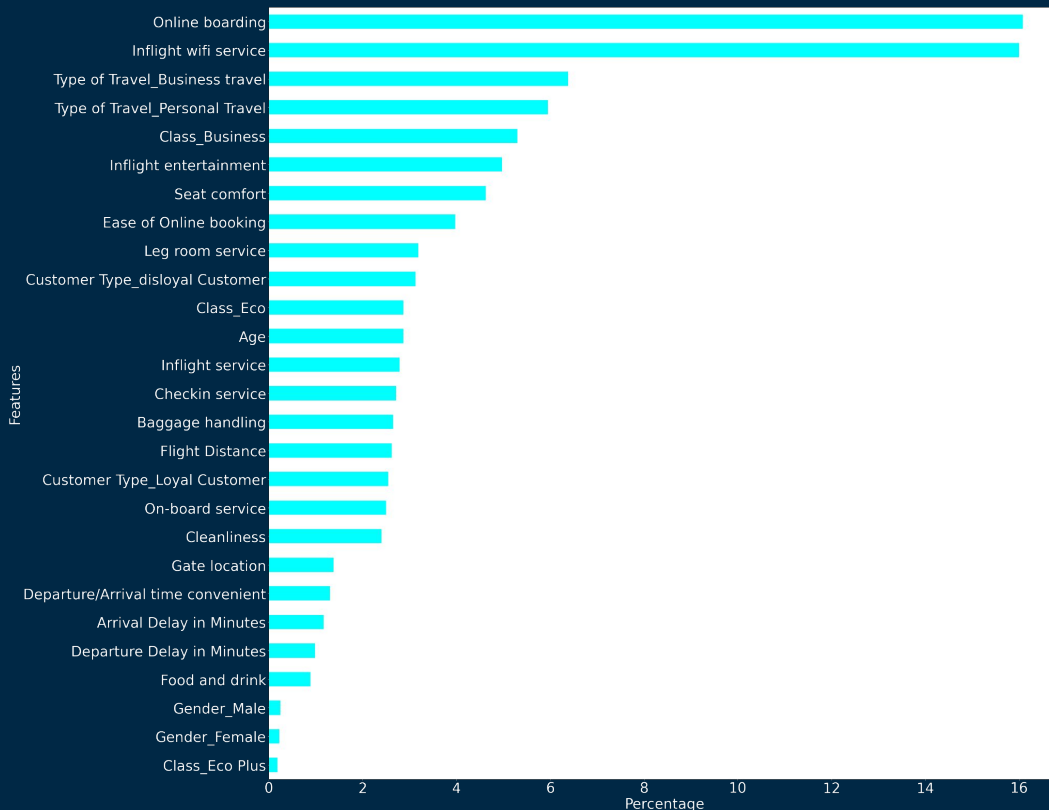




# Feature Importance

Possiamo notare che 'Online Boarding' e 'Inflight Wifi Service' sono le principali caratteristiche per la risoluzione della classificazione. Purtroppo nessuna feature è davvero preponderante nell'aiutarci a risolvere il problema. Sembra ovvio che siano fattori diversi a determinare la soddisfazione dei clienti in quanto individui diversi con bisogni diversi.

È interessante notare che la metrica riguardante la fedeltà di un cliente è inversamente proporzionale all'influenza sulla soddisfazione.





# Conclusioni

Sulla base dei nostri risultati, consigliamo alla compagnia aerea di prestare particolare attenzione all' **Online Boarding**. Tenzionalmente porta alla non soddisfazione dei clienti e potrebbe essere opportuno chiedere ulteriori feedback per come migliorarlo.

Il **servizio Wifi a bordo** aiuta a predire in maniera piuttosto chiara la soddisfazione dei clienti. Si potrebbe pensare di migliorarlo nel caso alcuni voli abbiano un segnale insufficiente, anche per avere un vantaggio competitivo.

Inoltre consigliamo di prestare attenzione alla classe **Business** in quanto i clienti non sembrano apprezzare particolarmente il servizio offerto.

In futuro, potrebbe essere interessante esaminare ulteriori fattori che influenzano la soddisfazione dei clienti, come l'orario del volo o il tempo d'attesa dei bagagli una volta atterrati. Inoltre, potrebbe essere utile implementare un sistema di feedback in tempo reale per raccogliere ulteriori informazioni sulla soddisfazione dei clienti e migliorare il servizio offerto dalla compagnia aerea.