

Descrizione dei campi

- SERVICETYPE: tipo di servizio/mezzo utilizzato per il viaggio
- VEHICLETYPE: tipo di veicolo
- DEPARTURE_COUNTRY: nazione di partenza
- DEPARTURE_ZIPCODE zipocde (CAP) di partenza
- ARRIVAL_COUNTRY: nazione di arrivo
- ARRIVAL_ZIPCODE: zipcode (CAP) di arrivo
- SHIPPING_DATE data di spedizione
- GROSS_WEIGHT_KG: peso lordo
- NET_WEIGHT_KG: peso netto
- VOLUME_M3: volume
- DECLARED_DISTANCE_KM: distanza tra il punto di arrivo e quello di partenza
- ACTUAL_DELIVERY_DATE: data di arrivo
- DELIVERY_TIME_HH: tempo di viaggio
- WDAY: giorno della settimana della data di spedizione (0 = lunedì)

Osservazioni

- Si vede chiaramente che i viaggi che iniziano tra giovedì e domenica hanno durata più lunga; questo è probabilmente legato al fatto che i depositi sono chiusi nel weekend; come ci si deve regolare per gestire questo dato?
- C'è un mix di viaggi tra località italiane, statunitensi, e in qualche caso viaggi italia/USA; ha senso considerare tutti insieme? Meglio fare analisi separate o concentrarsi solo su un sottoinsieme?
- Alcune date sono palesemente errate, ci sono casi in cui la data di arrivo è antecedente alla data di consegna
- Alcune delle colonne contengono dati mancanti; ad esempio la distanza in km manca quasi sempre, ed è ovviamente un dato decisivo; è possibile ricostruirla a partire dagli ZIP code, ma richiede una elaborazione extra; noi lo abbiamo fatto per i codici italiani; per quelli US (mancanti) potrebbe essere interessante usare questo: <https://github.com/snkinitin/us-zipcode-distance> (per i viaggi IT/USA... non c'è modo, ma si tratta di viaggi ben diversi)
- I dati fanno riferimento a inizio anno, le osservazioni fatte ci permettono di estendere le stesse considerazioni per eventuali previsioni/realizzazione di modelli di ML? In generale no, la stagionalità è di solito una considerazione importante, e fare estrapolazioni al di fuori del periodo dell'anno considerato potrebbe portare ad errori