**Key Points presentazione Connected Cars**

**Slide 2 – Executive Summary**

**Slide 3 - Agenda**

**Slide 4 – Contenuti**

**Note:** I record/osservazioni sono le righe del dataset.

**Slide 5 – Metodologia**

**Obiettivo:** Illustrare il processo tramite il quale sono stati ottenuti i findings. Una descrizione dettagliata della metodologia è disponibile nell’allegato tecnico.

**Note:** I “+” rappresentano punti di valore aggiunto maggiore.

**Slide 6 – Distribuzione geografica**

**Note:** Per ogni vettura è stata identificata la timezone in cui si hanno più osservazioni. La dimensione dei punti indica il numero di vetture per ogni timezone.

**Slide 7/8 – Facts and figures**

**Obiettivo:** descrivere informazioni di base sull’uso delle vetture.

**Note:** Sull’asse y dei grafici si trova il numero di auto, sull’asse x i valori della variabile in questione. Quindi, prendendo ad esempio il primo grafico, quasi 300 auto sono state utilizzate tra lo 0 e il 10% dei giorni.

I giorni di utilizzo sono stati calcolati come frazione dei giorni di utilizzo a velocità maggiori di zero su i giorni trascorsi tra la prima osservazione del dataset e il 29 novembre.

Le velocità massime sono state calcolate come la media delle velocità massime giornaliere di ogni auto.

I km adesso rappresentano i km percorsi in media al giorno.

Le ore di guida al giorno fanno riferimento esclusivamente alle serie che l’algoritmo ha ritenuto utili all’analisi, escludendo quindi i piccoli spostamenti poco rilevanti.

**Slide 9/10 – Facts and figures**

**Obiettivo:** confrontare il comportamento degli utenti europei con quello degli utenti americani.

**Note:** per “America” si intende il continente, quindi Canada, Stati Uniti e Messico. I grafici mostrano la distribuzione delle auto in America in alto e quella delle auto in Europa in basso.

I comportamenti dei due gruppi di utenti sono simili, ma

1. In proporzione le auto americane tendono ad essere usate più spesso
2. Per quanto riguarda le velocità, tra le auto americane i valori agli estremi sono più rari: ad esempio, ci sono poche auto con velocità massime inferiori agli 80 km/h o superiori ai 220 km/h.
3. I km giornalieri medi sono simili per le due distribuzioni, con l’importante eccezione delle auto che hanno percorso meno di 25 km al giorno, che in Europa sono molto numerose.
4. La distribuzione delle ore guidate in media al giorno è simile, ma gli utenti americani che guidano meno di 15 minuti al giorno o più di 2 ore al giorno sono in proporzione più rari.

**Slide 11 – Cluster analysis**

**Note:** per ulteriori informazioni sul funzionamento dell’analisi dei cluster e sulla scelta del numero di cluster vedere l’allegato metodologico.

Le variabili utilizzate per il clustering sono:

1. Percentuale di giorni in cui è stata utilizzata l’auto
2. Media delle velocità massime giornaliere
3. Velocità media
4. Chilometri totali percorsi dalla vettura
5. Durata di guida media giornaliera.

**Slide 12 – Findings**

**Obiettivo:** presentare l’esistenza di cinque cluster.

**Note:** Per l’analisi dei cluster sono state analizzate tutte le auto non di test per le quali nel dataset si hanno almeno tre giorni di record a velocità superiori a zero, in totale 1179 auto.

Sul grafico gli assi rappresentano molteplici variabili, l’obiettivo è semplicemente quello di vedere i 5 cluster.

**Slide 13 – Findings: data**

**Obiettivo:** mostrare i valori tipici di ogni cluster per quanto riguarda delle variabili base.

**Note:** Il valore in grassetto è la media per quanto riguarda la variabile dei valori delle auto del cluster. I valori tra parentesi indicano il minimo e il massimo.

In questo caso i km rappresentano i km totali percorsi dalla vettura.

**Slide 14 – Findings: interpretation**

**Obiettivo:** dare un’interpretazione ai valori della tabella della slide 13.

**Slide 15 – Findings**

**Obiettivo:** mostrare l’appartenenza ai cluster per continente di appartenenza delle auto.

**Note:**

La percentuale di auto non utilizzate è simile nei due paesi, le auto di test sono meno in America. I *cluster 1 – Utilizzo semi-quotidiano* e *4 – guida nel weekend* hanno un’importanza maggiore in America. Il *cluster 3 – Piloti professionisti* mantiene le stesse dimensioni. L’appartenenza ai *cluster 2 – guida tranquilla* *e 5 – guida in pista* è più rara in America. Questo è dovuto alle differenze tra continenti viste nelle slide 9/10.

**Slide 16 – Findings**

**Obiettivo:** mostrare graficamente i valori illustrati nella tabella della slide 13.

**Note:** I grafici rappresentano la distribuzione delle vetture, divise per cluster, rispetto alle variabili già viste nelle slide 7 e 8. L’asse y rappresenta il numero di auto, l’asse x i valori della variabile.

**Slide 17 – Findings**

**Note:** La mappa sulla sinistra mostra il numero di auto rilevate all’interno delle coordinate di 29 autodromi italiani. La dimensione dei punti dipende dal numero di auto rilevato all’interno delle coordinate.

Nella mappa a destra ogni punto rappresenta un’osservazione del dataset alla latitudine e longitudine indicate sugli assi. I punti in nero rappresentano vetture di test oppure vetture che sono state utilizzate meno di tre giorni e quindi non hanno un cluster di appartenenza.

In totale sono rappresentate 28 auto:

1. 1 del *cluster 1 – utilizzo semi-quotidiano*
2. 1 del *cluster 2 – guida tranquilla*
3. 7 del *cluster 3 – pilota professionista*
4. 3 del *cluster 4 – guida nel weekend*
5. 15 del *cluster 5 – guida in pista*
6. 2 senza cluster

**Slide 18 – Findings**

**Obiettivo:** Mostrare quanto spesso le tre modalità di guida vengono utilizzate dalle vetture dei cluster.

**Note:** I grafici mostrano in che percentuale dei giorni rispetto al numero di giorni di utilizzo le tre modalità di guida sono state utilizzate. L’asse y mostra il numero di auto.

**Slide 19 - Findings**

**Obiettivo:** spiegare la lettura delle box plots.

**Note:** La linea centrale rappresenta la mediana. Entro la “scatola” che la circonda (interquartile range) si trovano il 50% delle osservazioni della distribuzione. 25% di osservazioni si trovano nella riga al di sopra della “scatola”, il restante 25% nella riga al di sotto. I punti che si trovano al di fuori di queste linee (outliers) fanno parte del 25% di osservazioni delle righe, ma si trovano a più di 1.5 volte sopra o sotto l’altezza della “scatola”.

**Slide 20 – Findings**

**Obiettivo:** Mostrare come l’uso dell’active launch control è estremamente raro, viene utilizzato occasionalmente, ma sempre molto raramente (tendenzialmente meno dello 0.05% dei giorni), da utenti del *cluster 3 – pilota professionista*.

**Slide 21 – Possibili applicazioni**

**Slide 22– Possibili applicazioni**

**Obiettivo:** mostrare come l’analisi svolta può essere utilizzata per ulteriori applicazioni, come l’elettrificazione del motore.

**Note:** per le slide successive la distanza minima perché due serie vengano considerate distinte è stata impostata a due ore. Gli spostamenti che hanno meno di un’ora di pausa tra di loro vengono considerati come un viaggio unico. Le slide rappresentano esclusivamente auto europee. Sono state incluse nel conteggio delle serie anche quelle considerate poco utili all’analisi dall’algoritmo.

**Slide 23 – Possibili applicazioni**

**Slide 24 – Possibili applicazioni**

**Note:** le ore di guida in questo caso potrebbero essere superiori a quelle viste nelle slide all’inizio della presentazione. Questo avviene perché all’interno di questo tempo sono contate anche le pause inferiori a due ore, cosa che invece non avveniva nel resto della presentazione.

Le ore sono espresse in base 100, quindi 1.25 = 1h15m, 1.5 = 1h30m, 1.75 = 1h45m.

**Slide 25 – Possibili applicazioni**

**Note:** il grafico mostra la distanza temporale media tra due viaggi nel caso ne avvengano almeno due nel corso dello stesso giorno.

**Slide 26 – Possibili applicazioni:**

**Note:** le ore di guida in questo caso potrebbero essere superiori a quelle viste nelle slide all’inizio della presentazione. Questo avviene perché all’interno di questo tempo sono contate anche le pause inferiori a due ore, cosa che invece non avveniva nel resto della presentazione.

**Slide 27 – Possibili applicazioni**

**Slide 28 – Possibili applicazioni**

**Obiettivo:** Mostrare la distribuzione delle auto nei cluster sulla base del continente di appartenenza. Le informazioni trasmesse dalla slide sono essenzialmente le stesse dei grafici a torta.

**Note:** Il grafico rappresenta il clustering come nella slide 12. L’unica differenza è che i punti, che rappresentano le vetture, hanno una forma piuttosto che un’altra sulla base del continente di appartenenza.

**Slide 29 – Possibili applicazioni**

**Obiettivo:** Mostrare la distribuzione delle auto nei cluster sulla base del continente di appartenenza.

**Note:** Il grafici rappresentano il clustering come nella slide precedente. I due grafici contengono rispettivamente solo le auto europee e solo quelle americane.