EXPLORATORY DATA ANALYSIS

WAYMO

Open Dataset

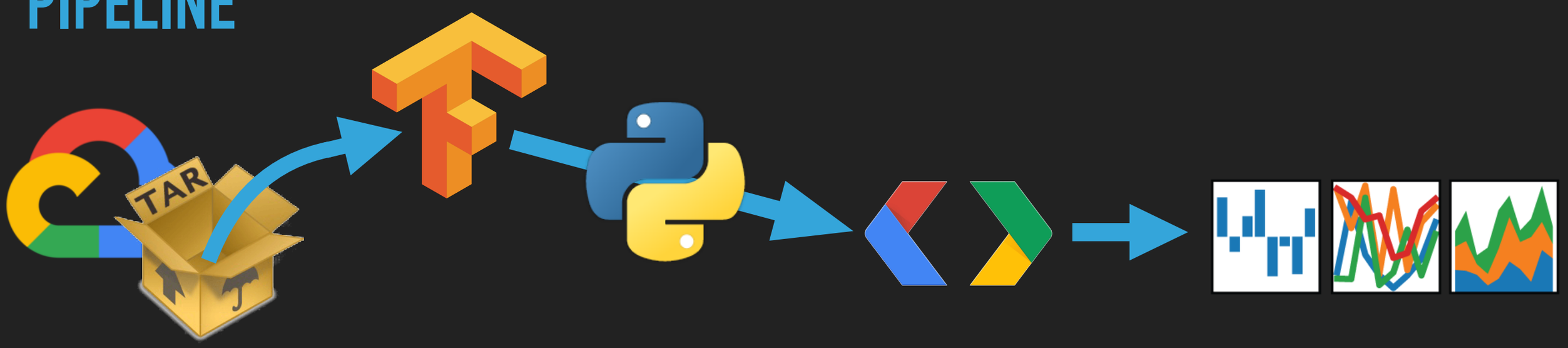
KATERINA SCHULZ | GALVANIZE DATA SCIENCE IMMERSIVE

SF, CA 2019

OVERVIEW

- Autonomous Vehicle development company, subsidiary of Alphabet
- Waymo cars collect LIDAR (“3D”) and camera (“2D”) data
- LIDAR measures distance to a target by laser; used for 3D scene reconstruction

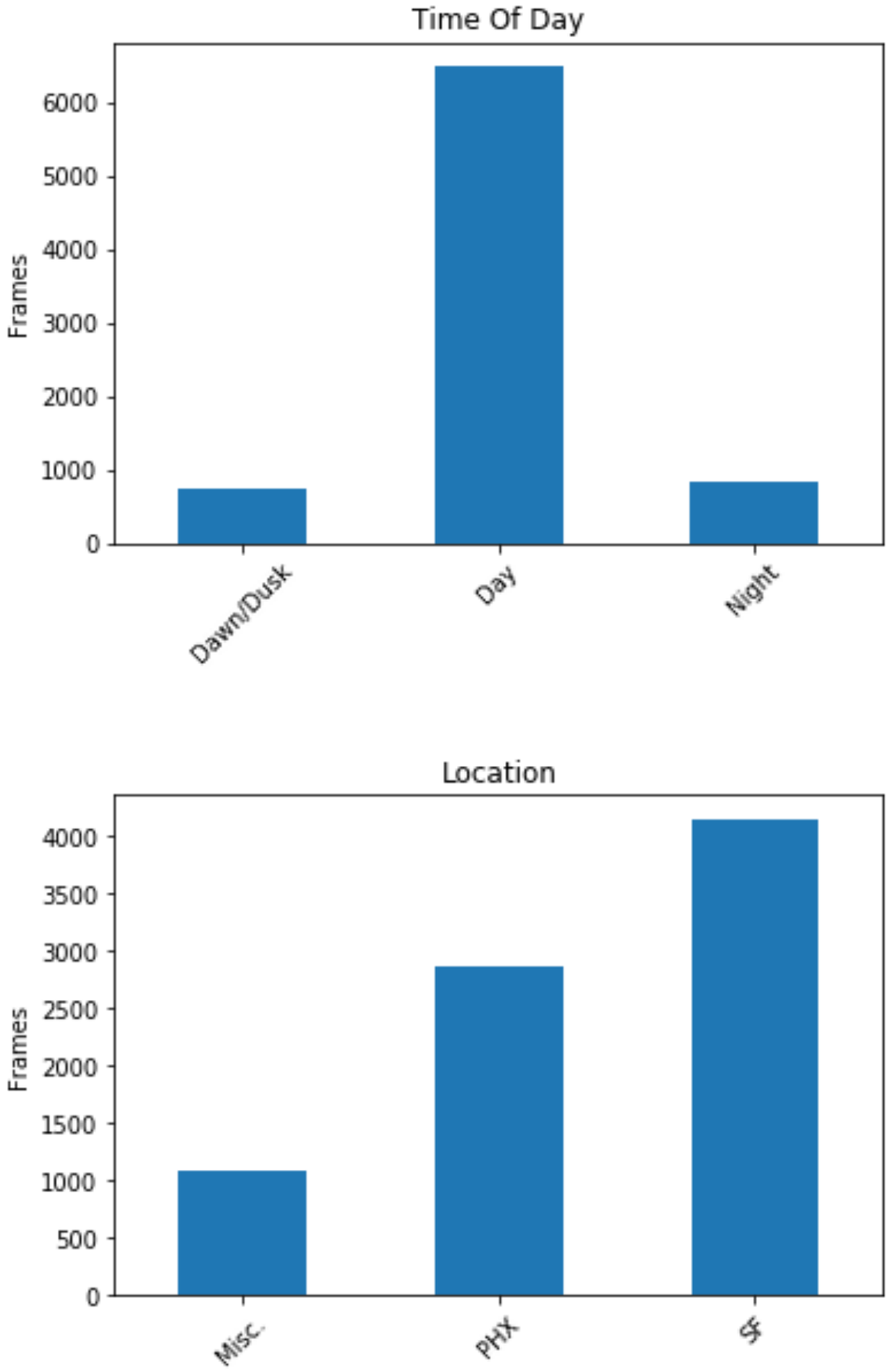
PIPELINE



Cloud → Unpack Tarballs → Extract Proto Buffs from TF Records → Write data to Pandas DF

SCENE ATTRIBUTES

- Representative sample of 8,000 frames
- 85% scenes located in SF, Phoenix
- 80% daytime scenes
- 99% “Sunny” weather



SCENE ATTRIBUTES: OBJECT CLASS COUNTS

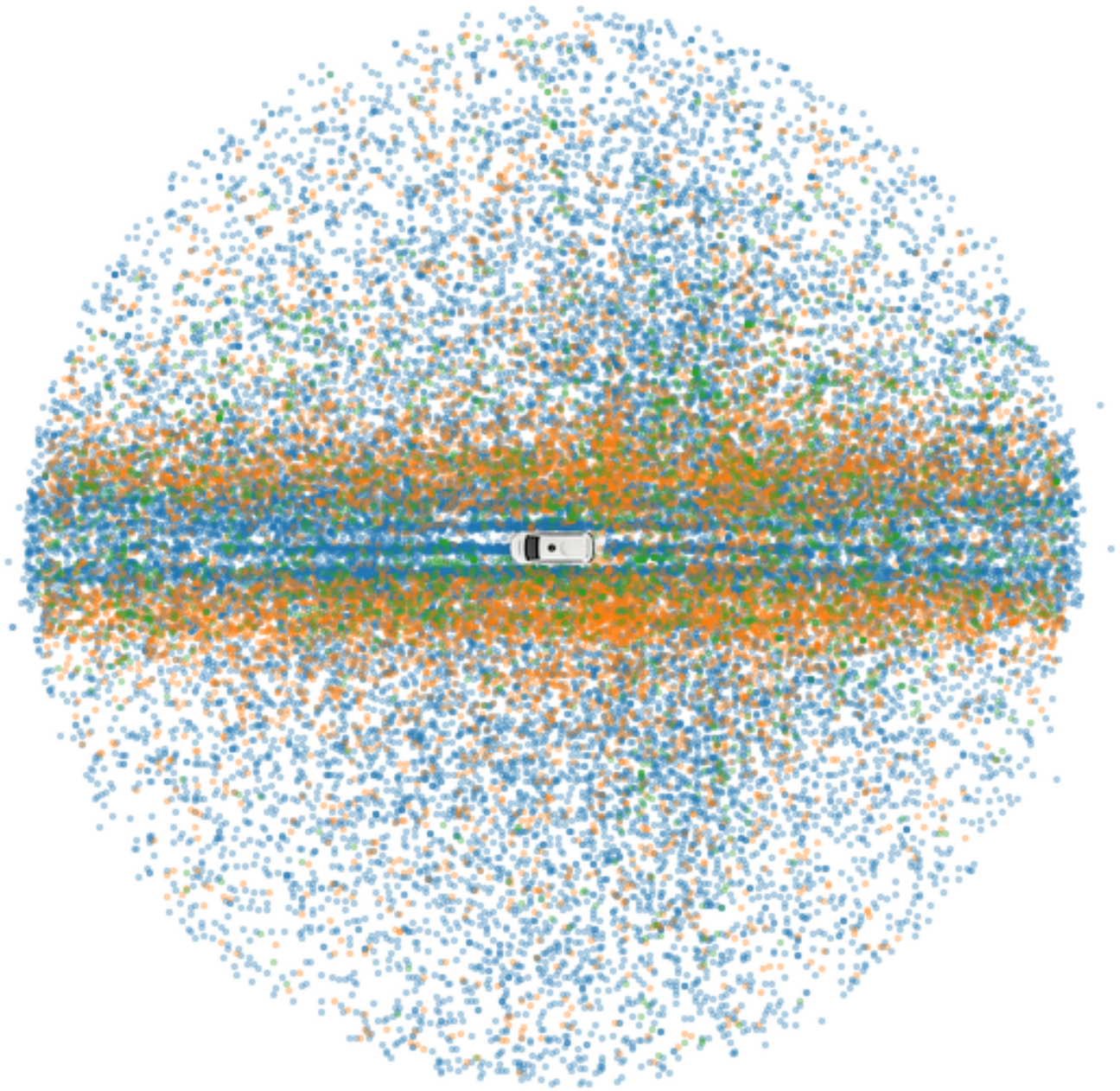
Object Class	Mean	Median	Max
Vehicle	30	27	163
Pedestrian	14	27	192
Cyclist	0	0	11
All	45	33	234

- Vehicle and Pedestrian classes make up majority of “mobile” object instances

OBJECT INSTANCES
SCATTER PLOT

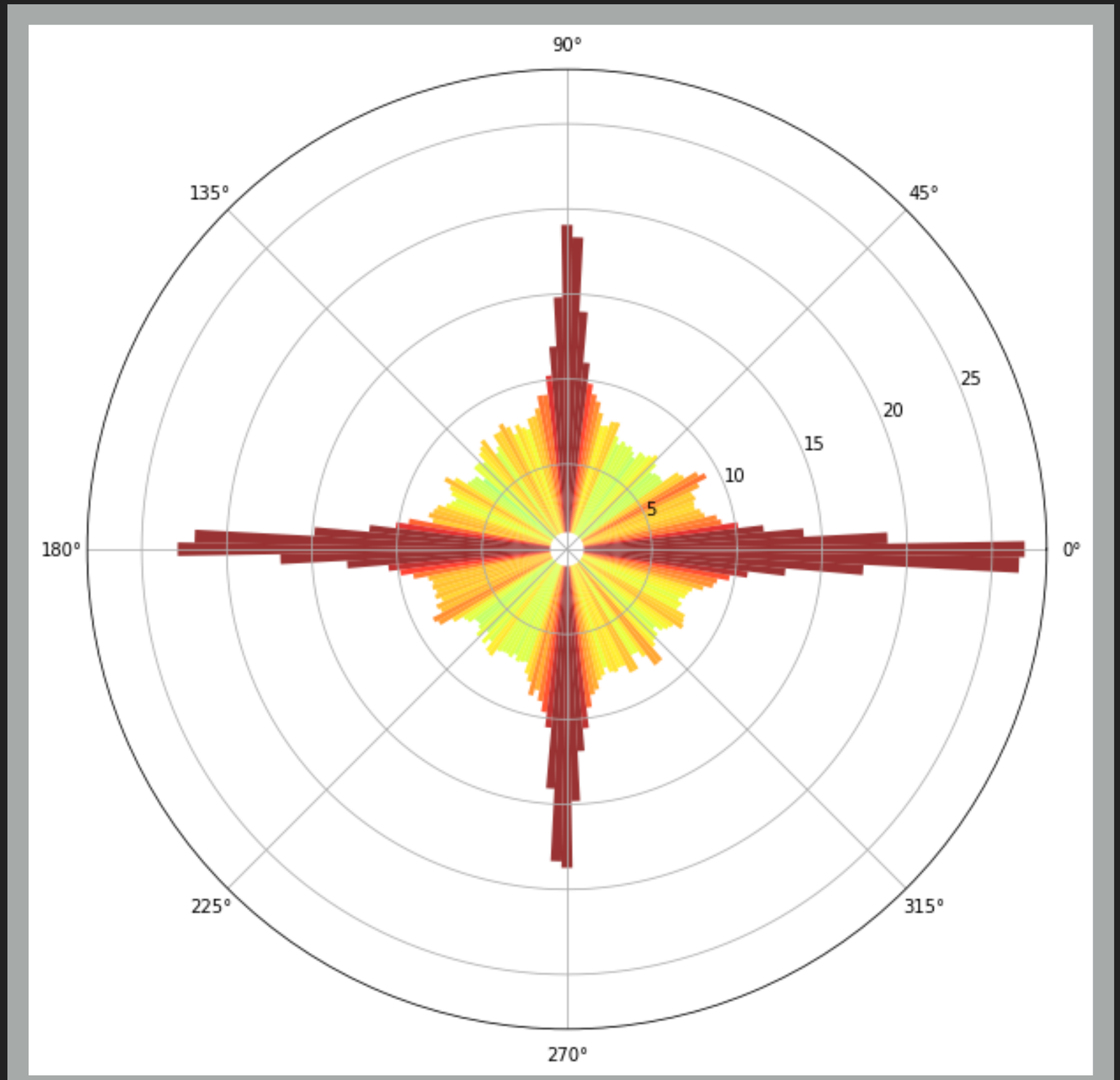
- Vehicle Class
- Pedestrian Class
- Cycle Class
- Waymo Car

- 38,000 Instances:
  - 25,000 Vehicle
  - 10,000 Pedestrian
  - 2,629 Cycle



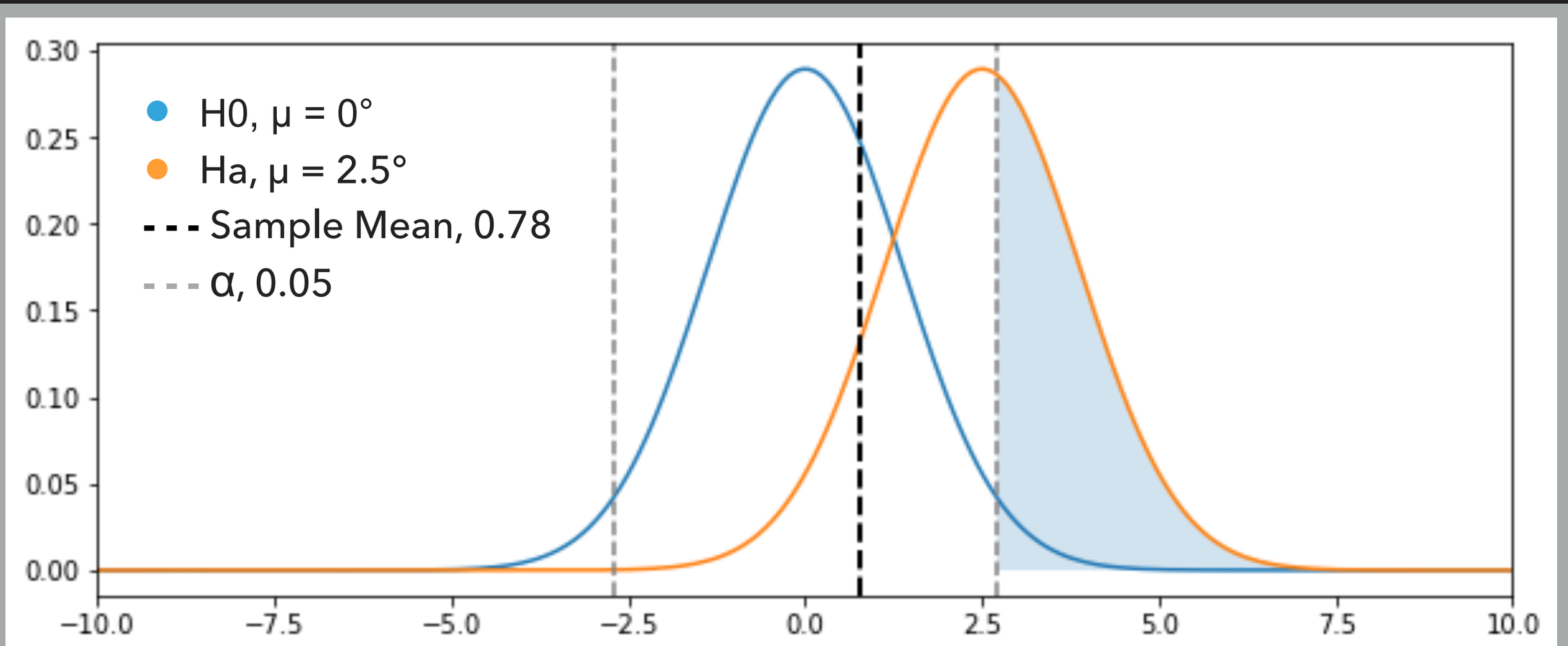
VEHICLE HEADINGS
HISTOGRAM

- Heading Parallel to Waymo Car at 0° and 180°
- Heading Orthogonal to Waymo Car at 90° and 270°
- 149,900 of 245,218 within 1° of Parallel or Orthogonal Heading



VEHICLE HEADINGS
HYPOTHESIS TESTING

- H_0: $\mu_{Heading} = 0°$
- H_a: $\mu_{Heading} \neq 0°$
- 95% CI: [-1.9, 3.5]
- Power: 0.664



QUESTIONS

waymo.com/open

github.com/kittyschulz/EDA-Waymo-Open-Dataset