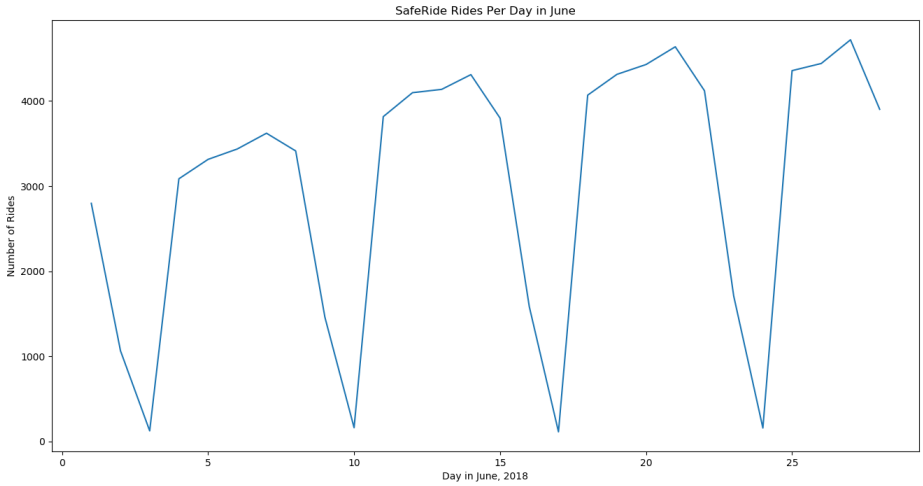
Mapping and Predicting Ride Demand for SafeRide™

Background and Problem Statement:

SafeRide is a non-emergency medical transport app that provides approximately 3000 rides per day to and from medical services locations (hospitals, doctors' offices). They partner with hospital networks and medical transport suppliers in Portland, OR, Kansas City, MO, Santa Monica, CA, and Chicago, IL. As a growing company, SafeRide has focused most of its energy on the usability of its interface and has done relatively little work to understand trends in its data.In order to best align the supply of medical transport vehicles to the demand for them, Saferide requests a geographic and time-based analysis of demand for their services. They also are interested in using a predictive model of where demand will occur in order to best position their partners' supply of medical transport vehicles.

Data:

As different markets have different patterns in demand and SafeRide works in very different markets, we will be using data on roughly 100,000 rides over the course of a month from the Portland market with these features:



- Pickup and dropoff location (latitude, longitude & address)
- Patient ID, Driver ID
- Ride type (ambulance van or wheelchair van)
- Timestamps including time ride sent to driver, time of pickup, time of dropoff
- Ride cancellation and reason (along with some data about the medical appointment)

See appendix for all features. To the right you can find an initial visualization of the number of rides per day in June by SafeRide in Portland.

Approach:

MVP for this capstone will include:

- Mapping of ride demand (point maps and heat maps) using Folium for peak demand and daily demand.
- EDA for ride demand and provision including peak demand time, peak demand time locations, customer time spent waiting after requesting ride
- Time-series-based prediction of demand overall using Boosting, Random Forest, and Linear Regression algorithms and testing of that prediction on train-test split (possibly using Neural Networks)

MVP+ for this capstone would include:

- Time-series-based prediction of rides in individual geo-hashes (size still TBD) using above methodology.
- Predicting ride cancellation, based on customer appointment characteristics and previous cancellations.

Building for Next Capstone:

It is my aim to build ride demand predictions so that in the next capstone I can suggest optimal numbers of drivers to have available and (MVP+) propose optimal locations of their drivers to be able to quickly serve the present demand.

Appendix: Fields in Data Extract

Field Name	Rename Field As
rideId	rideId
fromAddress	fromAddress
toAddress	toAddress
distance	distance
rideType	rideType
vehicleCompanyName	vehicleCompanyName
requestedVehicleType	requestedVehicleType
vehicleType	vehicleType
vehiclePlateNumber	vehiclePlateNumber
pickupTime	pickupTime
appointmentTime	appointmentTime
finalCost	finalCost
estimatedCost	estimatedCost
passengerCost	passengerCost
careCost	careCost
selfPay	selfPay
creditCardFee	creditCardFee
saferideFee	saferideFee
status	status
driverName	driverName
bookedBy	bookedBy
fromLatitude	fromLatitude
fromLongitude	fromLongitude
toLatitude	toLatitude
toLogitude	toLogitude
rideSentToDriver	rideSentToDriver
rideNotes	rideNotes
createdDate	createdDate
updatedDate	updatedDate
acceptedDate	acceptedDate
startedDate	startedDate
arrivedDate	arrivedDate
completedDate	completedDate
cancelledDate	cancelledDate
cancelledBy	cancelledBy
cancelledReasonType	cancelledReasonType
cancelledReasonMessage	cancelledReasonMessage
hospitalName	hospitalName
hospitalPhoneNumber	hospitalPhoneNumber
patientMedicalId	patientMedicalId
patientFirstName	patientFirstName
patientLastName	patientLastName
patientPhoneNumber	patientPhoneNumber
patientNotes	patientNotes
lyftRideId	
lyftReferenceCode	
vehicleCompanyManaged	vehicleCompanyManaged
bulkPurchased	bulkPurchased
RideRequestReason	RideRequestReason