Problem: Point clouds are an important data set consisting of points with XYZ coordinates. Point clouds are seen as an integral piece of the adoption of robotics into commercial life because of the ease of depth perception they provide. Transformers have gained popularity in many different areas of machine learning but have not yet been widely explored with point clouds.

Goal: The 3DETR architecture is a transformer point cloud detection network. The goal of this work is to extend the 3DETR architecture to a data set which is more representative of point clouds found in practice.

Data Set
The 3DETR architecture was originally trained on the Scannet and SUN RGB-D data sets which are both collected from 3D camera sensors. This means they do not provide point clouds, rather they are point sampled 3D meshes. This problem introduces the KITTI data set which is an outdoor point cloud data set collected from a Velodyne lidar scanner, making it more representative of point clouds found in practice.

Conclusion
Over the course of the REU we were able to extend the 3DETR architecture to the more realistic KITTI data set. As the results show, the network appears to approach a converging value for the loss and the mean average precision increases throughout the training indicating that the network is learning to detect the objects. The loss graph shows a non flat convergence line indicating that the network can be trained further to obtain better results on the mean average precision.