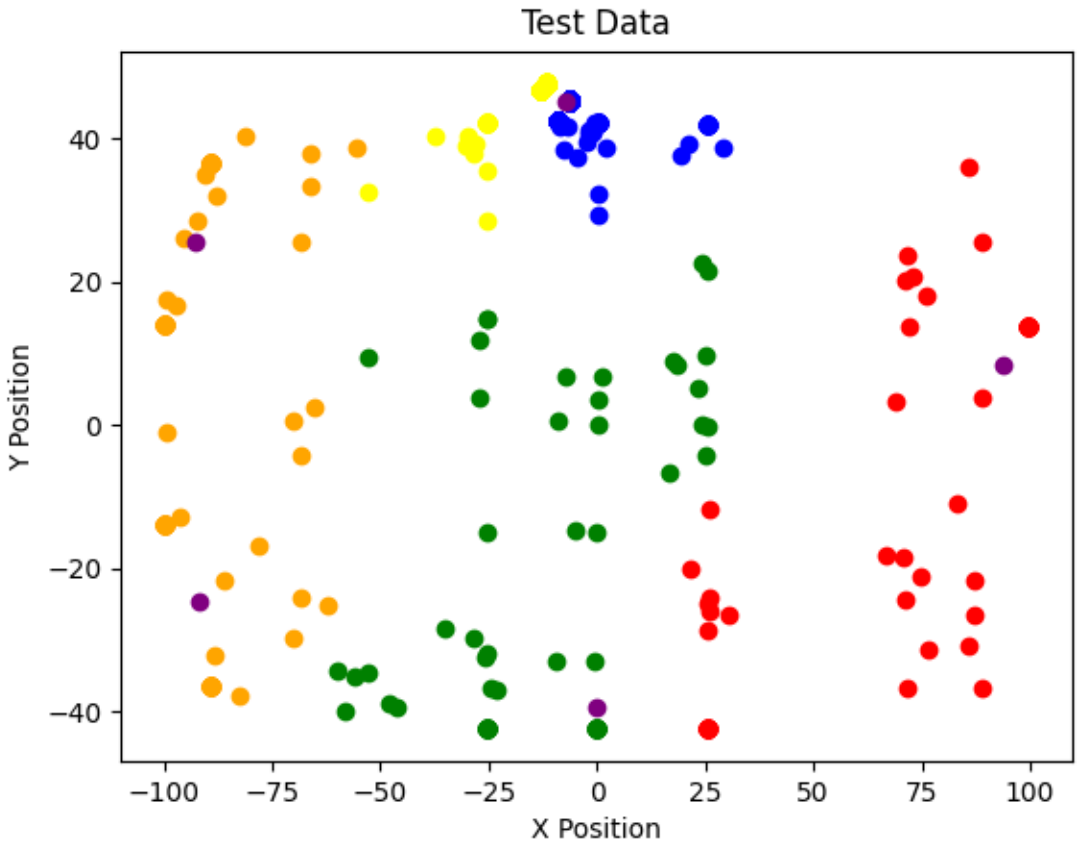
Trevor Davidson
SMT Data Science

For questions #1 and #2, a clustering algorithm was used to summarize the location gaps in the data. To read more about this, click here: https://scikit-learn.org/stable/modules/clustering.html#mini-batch-kmeans. For a simplification of this analysis, z-position data was removed when showing graphs

Question #1: Find when venue testing resulted in data gaps of >300ms and identify clusters on ice
# of Gaps across all tests: 4182
Approximate Locations:
(-7.324, 45.281)
(-91.732, -24.726)
(-0.273, -39.5685)
(93.859, 8.512)
(-92.762, 25.664)
The major takeaway from this data is that most errors occur at the player benches and near the boards. There are some issues in center ice, but these appear much less frequently.



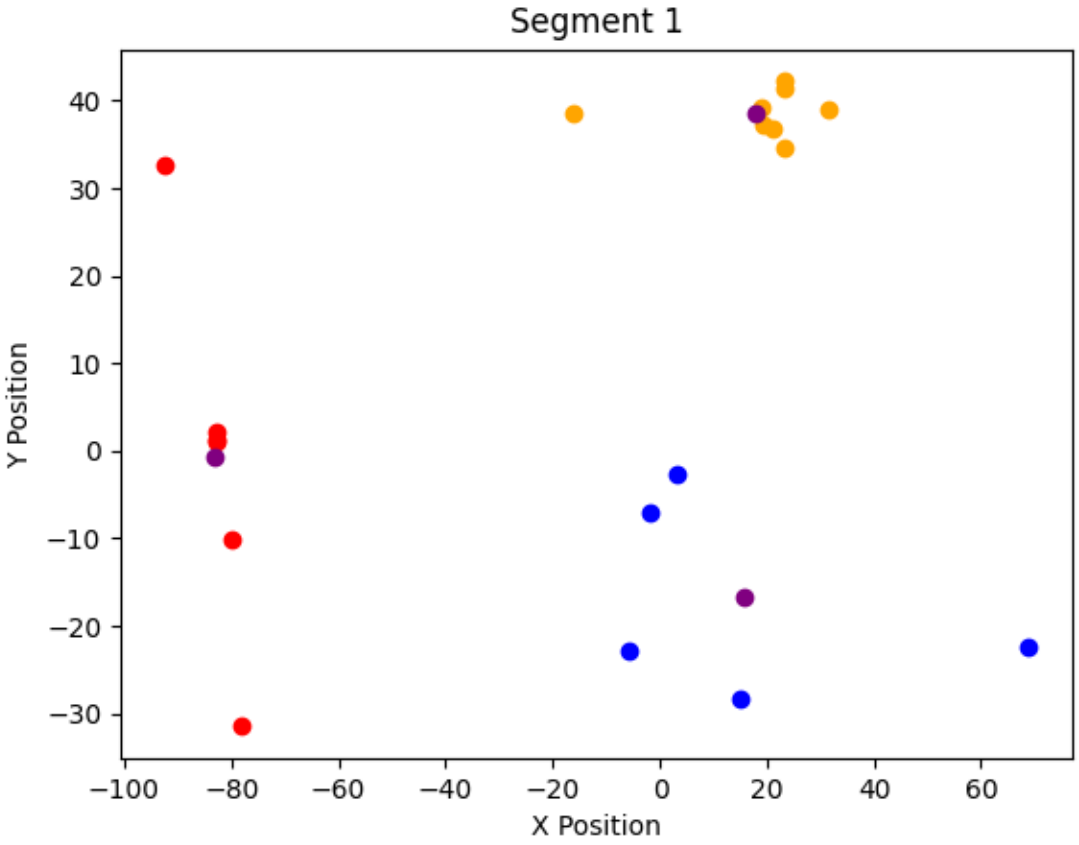
Trevor Davidson
SMT Data Science

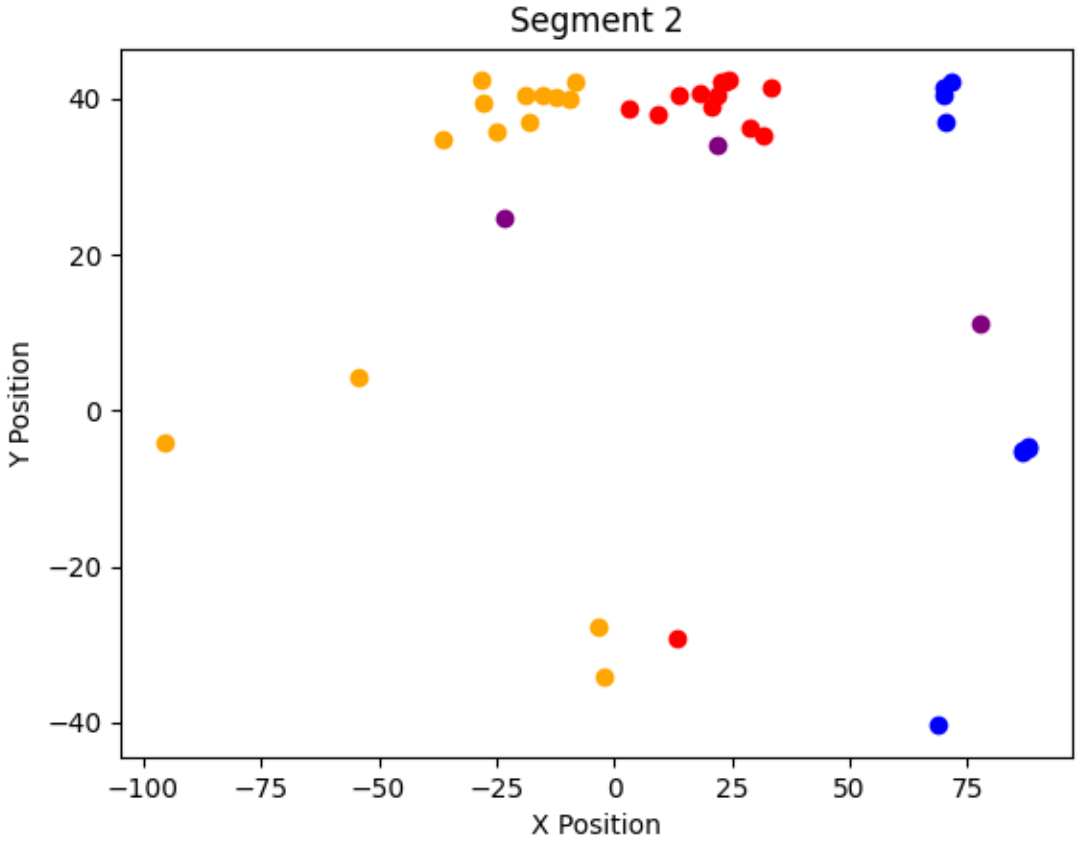
Question #2: Find >300ms data gaps for player segments, ignoring players on bench

Interval	# of Gaps	Locations (approximate)
1	19	(18.086, 38.629) (-83.215, -0.71) (15.921, -16.692)
2	36	(77.785, 11.183) (-23.499, 24.638) (21.928, 34.087)
3	54	(91.487, -5.413) (24.643, 39.836) (-41.348, 27.015) (-3.09, -24.375)
4	11	(86.768, 8.293) (-17.121, 33.864)

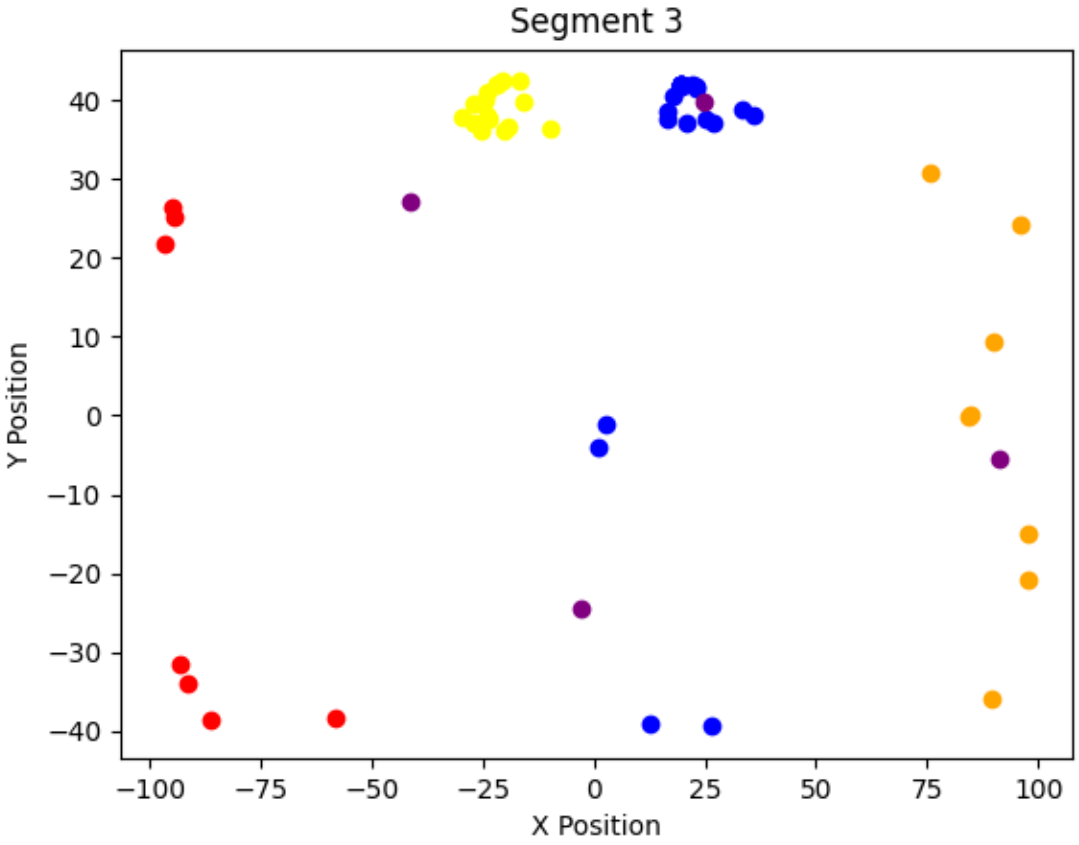
Eliminating the player bench locations was hugely beneficial between the counts in each section, with the number of errors dramatically decreasing. In particular, the graph of segment 2 displays how dramatically areas close to the player benches can cause significant issues in data collection relative to the rest of the ice. However, other areas of the ice that seem to be a problem are close to the boards. This is especially evident in the graphs of segment 3 and 4, where most error points occur within a few feet of the boards nearest to the player.

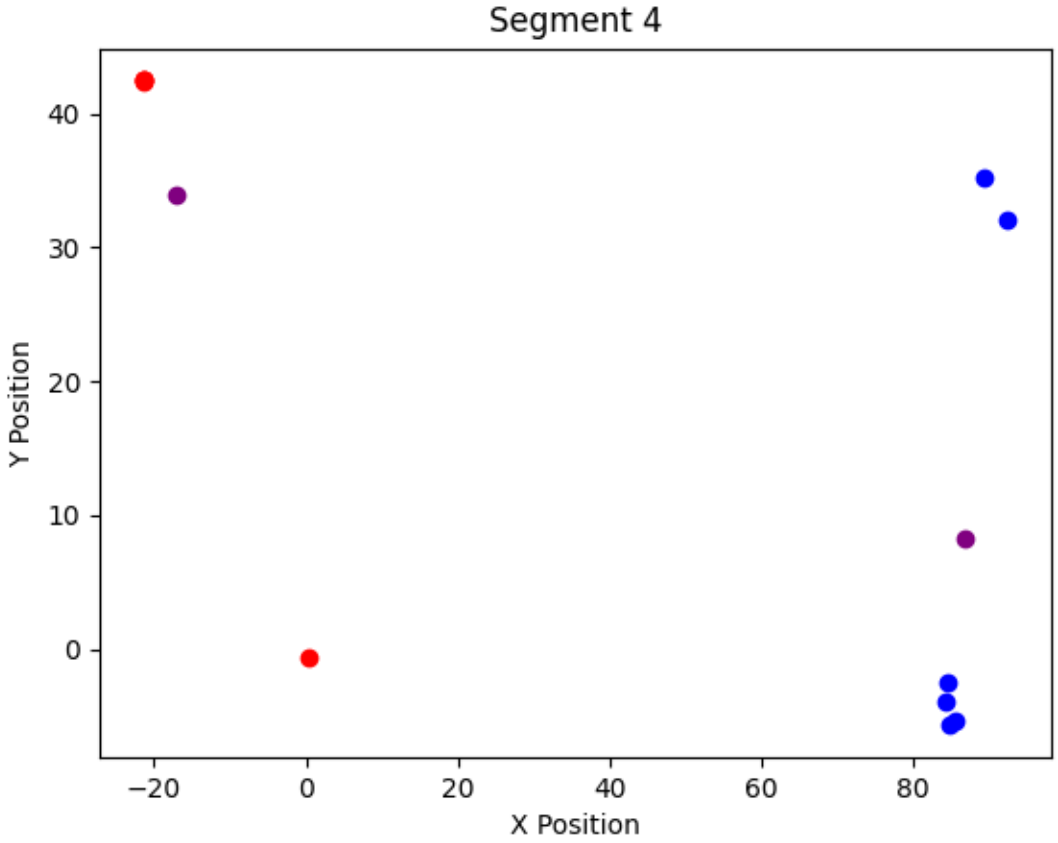
Trevor Davidson
SMT Data Science





Trevor Davidson
SMT Data Science





Trevor Davidson
SMT Data Science

For questions 3 and 4, Z-position data was removed, as the elevation of players does not change while on the ice. Additionally, the puck was removed from these calculations, since players are the focus of this analysis.

For questions 3 and 4, max speed was calculated using a rolling average of 10 data points. This was done to remove any jumps that may result in a max speed that was out of a reasonable range. However, it does sacrifice determining an instantaneous max speed that may have been greater than expected.

Question 3: Determine speed of the fastest player

Interval	Fastest Player	Max Speed (ft/s)	Location X	Location Y
1	208	27.729	-9.15	-16.11
2	207	27.614	-9.87	22.44
3	229	29.317	22.29	-31.38
4	119	29.182	-6.54	11.79

Question #4: Evaluate speed measurements for reasonable results

If there was an issue with the calculations done in question #3, one of the following things would be probably be true:

1) The maximum speed would be far too low (<10ft/s) to be an example of an NHL rush, or the maximum speed would be far too high (>40ft/s). A high data error could be the result of a data collection issue, electronics issue, or noise (combined with not using a rolling average)
2) The X location would be close to the end boards (within 20 feet). This is because players need to stop when going on the rush, players need space to accelerate to speed, or are battling down low for pucks. It's most likely that the maximum speed a player attains is in the neutral zone (-40 to 40), since that is before an offsides line and is usually when players begin to ready themselves for a pass or shot.
3) The Y location would be close to the side boards (within a few feet). This is because players generally do not get a lot of speed towards the edges of the ice, as they need space to take strides.
4) The direction of movement is in Y direction. Maximum speeds are usually attained on the rush, going towards the other team's goal. The Y direction also has much less space to attain speed, so any maximum would be much tougher to achieve moving in that direction. But, this should not be taken as rule, because it's possible a player crosses over to build speed. During that move, a player may try to turn while maintaining speed.

Based on these criteria, it seems unlikely that any of the data is an issue in this calculation.

Question #5: How would you find a genuine body check?

A body check is usually one player colliding with and transferring energy into an opposing player. There are a few things that could be used to determine this:

1) Two player tags in close vicinity to each other from opposite teams
2) A sharp decrease in velocity from at least one player in the vicinity

3) A sharp change in direction from at least one player in the vicinity. An exception to this could be a hit along the boards, but even then, the attacking player's rebound off the boards may result in a sharp reversal of direction
4) The nearby presence of the puck at a relatively recent time. This may only separate out legality of a hit, rather than whether a hit occurred, but considering most hits occur in the vicinity of the puck, this may be a good metric