

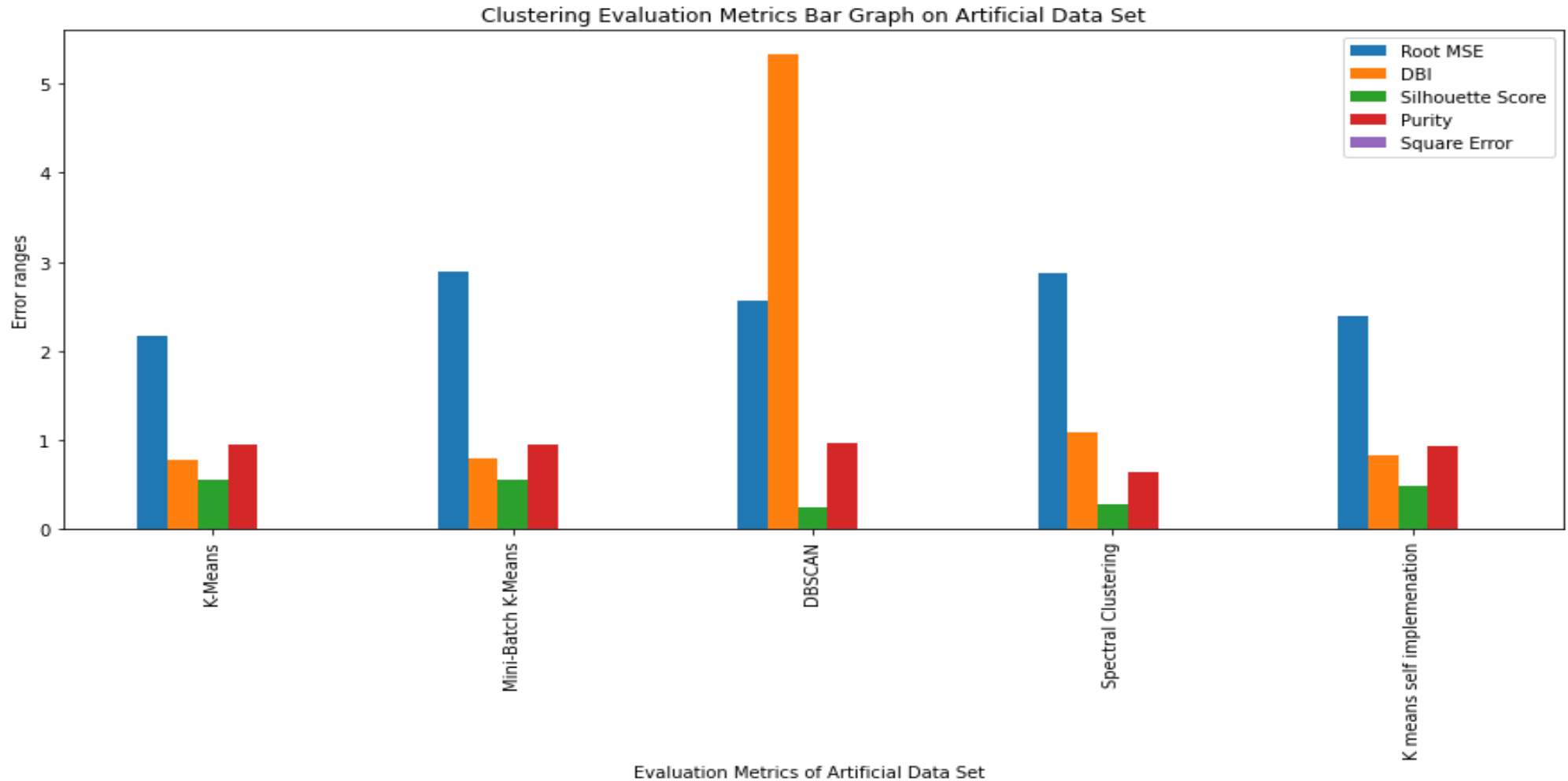
1 EVALUATION METRICS EXPLANATION

2 TABLE OF CONTENTS

1	Evaluation Metrics Explanation	1
3	Clustering Evaluation Metrics Bar Graph on Artificial Data Set:	2
3.1	Graph	2
3.2	Tables	3
3.3	Explanation	4
4	Clustering Evaluation Metrics Bar Graph on Artificial Data Set:	5
4.1	Graph	5
4.2	Tables	6
4.3	Explanation	6
5	Clustering Evaluation Metrics Bar Graph on Artificial Data Set:	7
5.1	Graph	7
5.2	Tables	8
5.3	Explanation	8

3 CLUSTERING EVALUATION METRICS BAR GRAPH ON ARTIFICIAL DATA SET:

3.1 GRAPH



3.2 TABLES

	Clustering Algorithm	Root MSE	DBI	Silhouette Score	Purity
0	K-Means	2.161395	0.774641	0.563799	0.957341
1	Mini-Batch K-Means	2.885720	0.792386	0.557206	0.959325
2	DBSCAN	2.559025	5.331623	0.255895	0.962302
3	Spectral Clustering	2.868134	1.089437	0.273916	0.648810
4	K means self implemenation	2.386304	0.828633	0.484632	0.930556

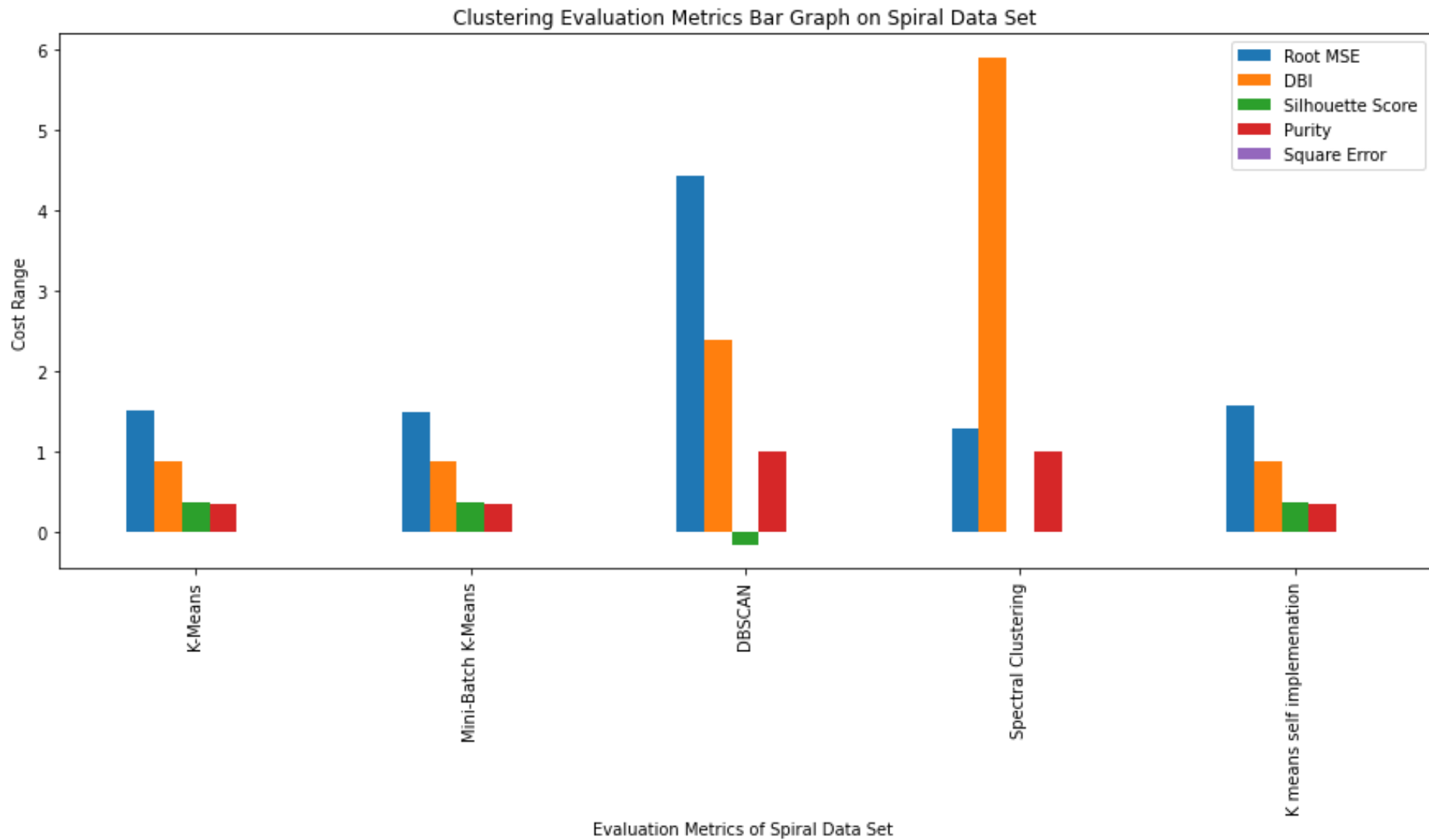
3.3 EXPLANATION

- 3.3.1 K mean and mini-batch k means are performing better in convex shape, this is convex shape data.
- 3.3.2 **Silhouette score:** of K Means and Minibatch K means is better than other algorithms. The silhouette score of 1 or close to 1 means that the clusters are very dense and nicely separated. Higher value is preferred
- 3.3.3 **DBI Score:** of K Means and Mini-batch K means is better than other algorithms. It is defined as a ratio between the cluster scatter and the cluster's separation and a lower value will mean that the clustering is better. Lower value is preferred.
- 3.3.4 But **spectral clustering and DB Scan** result is not good for convex shape it typically produces good result for non-convex shapes.
- 3.3.5 **Root mean Square Error:** Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the line of best fit.

Lower value is preferred. So RMSE of K means and mini-batch is lower than other algorithms.

4 CLUSTERING EVALUATION METRICS BAR GRAPH ON ARTIFICIAL DATA SET:

4.1 GRAPH



4.2 TABLES

	Clustering Algorithm	Root MSE	DBI	Silhouette Score	Purity	Square Error
0	K-Means	1.510946	0.887228	0.361942	0.347267	0.0
1	Mini-Batch K-Means	1.498123	0.886664	0.359947	0.350482	0.0
2	DBSCAN	4.436765	2.378795	-0.156053	0.996785	0.0
3	Spectral Clustering	1.743154	5.899294	0.000785	1.000000	0.0
4	K means self implemenation	1.563243	0.880736	0.360147	0.344051	0.0

4.3 EXPLANATION

4.3.1 DBSCAN and spectral clustering are performing better in non-convex shape, this is non-convex shape data.

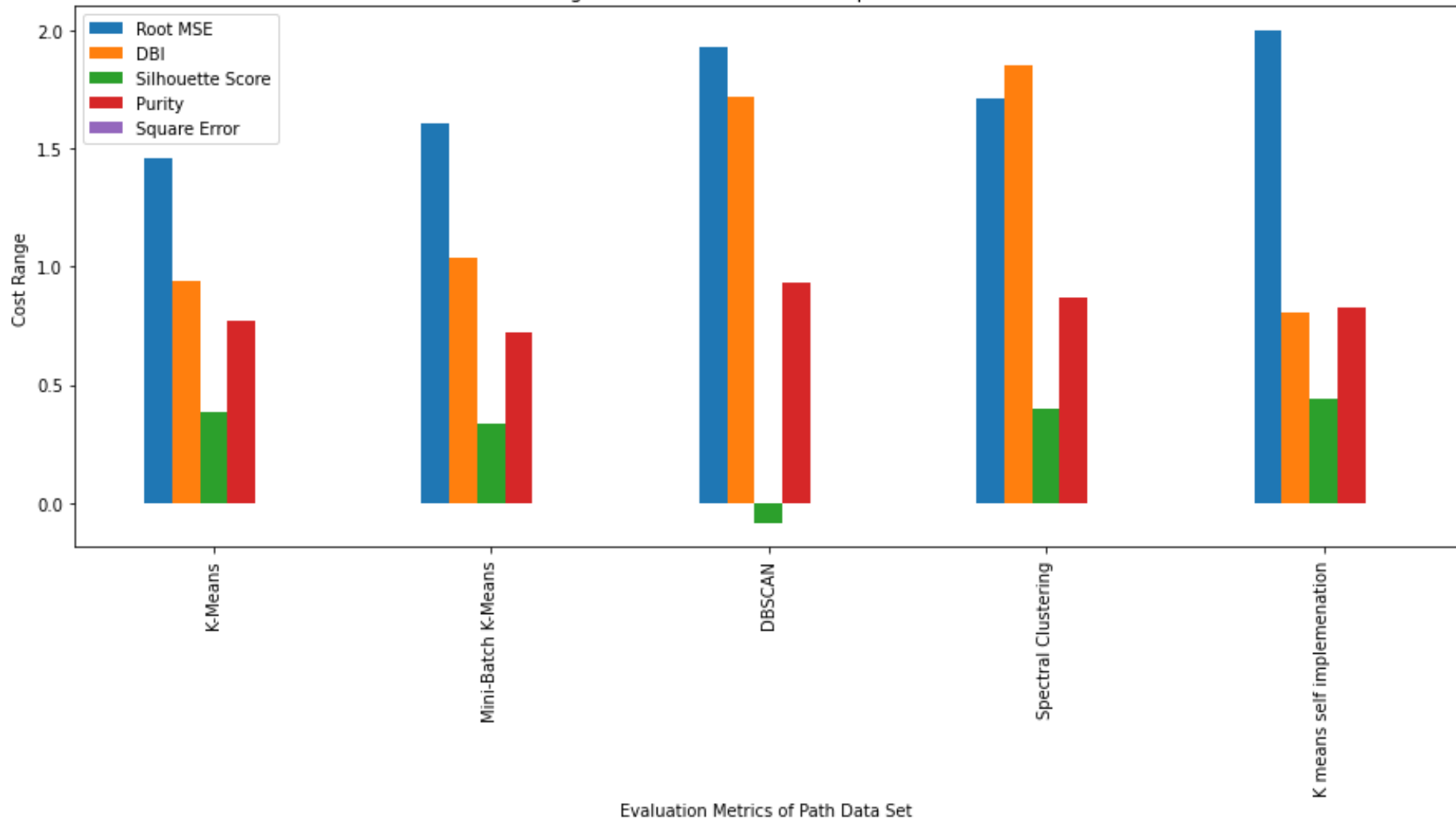
4.3.2 **Purity:** of K Means and Minibatch K means is better than other algorithms. Purity is a measure of the extent to which clusters contain a single class. If purity is 1 or close to one means that total number of objects(data points) that were classified correctly.

4.3.3 **Due to Non-Convex shape:** If We can evaluate the result only from purity then it will be better because its gave us accurate result but in other scores (DBI etc.) are not producing accurate score due to no-convex shape.

5 CLUSTERING EVALUATION METRICS BAR GRAPH ON ARTIFICIAL DATA SET:

5.1 GRAPH

Clustering Evaluation Metrics Bar Graph on Path Data Set



5.2 TABLES

	Clustering Algorithm	Root MSE	DBI	Silhouette Score	Purity
0	K-Means	1.461890	0.941059	0.384819	0.769231
1	Mini-Batch K-Means	1.607882	1.035558	0.332399	0.722408
2	DBSCAN	1.926752	1.718659	-0.083031	0.933110
3	Spectral Clustering	1.711656	1.849967	0.401102	0.869565
4	K means self implemenation	1.998327	0.806080	0.444112	0.829431

5.3 EXPLANATION

5.3.1 DBSCAN and spectral clustering are performing better in non-convex shape, this is non-convex shape data.

5.3.2 **Purity:** of DBSCAN and spectral clustering are better than other algorithms. Purity is a measure of the extent to which clusters contain a single class. If purity is 1 or close to one means that total number of objects (data points) that were classified correctly.

5.3.3 **Silhouette Score:** of Spectral Clustering is better because its performing better with non-convex shape