

No	Issues	Experiment	Draft	Date
1.	A new dataset for motivation? I have tried many times, could not find the best one.	-	-	3 Dec – 6 Dec
2.	Rectifying bound on Adaptive Pruning Schemes	Done	Not yet	-
3.	Applying multi queries shared computations <ul style="list-style-type: none"> - Combine target and reference query in one query instead of using two queries (done) - Combine aggregate function (Instead of views $(a_1, m_1, f_1), (a_1, m_2, f_2) \dots (a_1, m_k, f_k)$, each requiring execution of two queries, We combine these views into a single view $(a_1, \{m_1, m_2 \dots m_k\}, \{f_1, f_2 \dots f_k\})$) - Combine multiple GROUP BY (I don't know how to implement this) - Compare the performance among three scenarios (SeeDB with combine aggregate function, DiVE, DiVE with combine aggregate function) 	In progress	Not yet	3 Dec – 12 Dec
4.	Proposing another applicable distance for context driven similarity (currently use Jaccard) I have experiences using Cosine Similarity for measuring document similarity. I do not yet have a clear intuition on where one should be preferred over the other especially for the case of view similarity. Up to now I have no idea what another distance function that can be used except Jaccard and Cosine Similarity.	Need confirmation	-	-
5.	Dealing with hierarchical dataset. The dataset contains such as location dimensions (city, state, province, country), time dimensions (Year, Week, Quarter, days), product dimensions (product categories) commonly can be processed by OLAP queries. We need to find a way to calculate the similarity and dissimilarity between two views which have hierarchical attributes	In progress	Not yet	6 Dec – 13 Dec
6.	Prediction interval on adaptive Pruning. We decide to rely on non-parametric predictive interval models to determine maximum value with	Not yet	Not yet	10 – 14 Dec

	<p>certain level of confidence.</p> <p>Generally, PI can be defined as following:</p> <ul style="list-style-type: none"> • PI80: need to execute 9 sample of views • PI85: need to execute 12 sample of views • PI90: need to executes 20 sample of views • PI95: need to executes 40 sample of views • PI97: need to executes 60 sample of views <p>The result of our experiments show that PI can be used effectively on adaptive pruning. However, in our experiments, we used relatively small size of the number of views.</p> <p>I am not sure, if we have e.g., more than 10000 views that need to be evaluated. For instance, PI97 requires 60 sample of views, which is just very small size if we compare to the population (e.g., 10000 views)</p> <p>Is PI still applicable to this issue?</p>			
7.	<p>Axes Recommendations</p> <p>In our experiment, I defined by myself which attributes will be on X-axis and Y-axis.</p> <p>I put categorical attribute as X-axis and numerical attributes as Y-axis with the aggregate function.</p> <p>However, in some case, attribute such as 'age' more suitable to be X-axis with binning not Y-axis.</p> <p>How to deal with these kind of situation?</p>	Not yet	Not yet	10 – 14 Dec
8.	<p>False discovery</p> <p>I read some papers which focus on this issue. Up to now, we did not consider yet with the false discovery.</p> <p>For instance, one of view in top-k set does not make sense due to the wrong combination of axes.</p> <p>We did not consider the relationship between attributes to make sure that we are not discovering false views.</p> <p>In some works, they show that the view is statistically significant by applying hypothesis test.</p>	Not yet	Not yet	17 – 21 Dec

9.	<p>Coverage should be mentioned</p> <p>Two types of query load in the experiments:</p> <ol style="list-style-type: none"> 1. Compare between two subsets (e.g., disease vs. no disease) – <i>targeted/ we know what we want to do</i> (total number of views are less than No 2) 2. Compare between one subset to whole dataset (get all subsets from dataset then compare to whole dataset). For instance, Flights dataset has attribute = ‘carrier’ and there are a lot of carrier (e.g., AA, US, XX, UU), each carrier is a one subset. <p>If we only compare between each subset to whole dataset, it only show the trend of each subset compared to whole dataset.</p> <ol style="list-style-type: none"> 3. It seems interesting if we compare among subsets (e.g., AA vs. US, US vs. XX, AA vs. UU, etc). But it increases the number of combination significantly. <p>For instance, user may need to know the performance comparison of carrier AA vs. XX in terms of arrival delay. However, I think it depends on user want.</p> <p>Do we need to compare among all subsets in the dataset? or we need subset recommendation to know which subset that the user wants?</p> <p>In our experiment, we did a comparison between disease and no disease such as in no 1.</p> <p>If we run the experiment for the case of no 3, do we need to do the all combinations? or just sampling and mention the coverage?</p>	Not yet	Not yet	17 – 21 Dec
10.	<p>Subset recommendation</p> <p>This work will be the second topic of my PhD, which is focus on subset recommendation. I am going to used Active learning to select the sample subsets and include the user feedback.</p>	Not yet	Not yet	-
11.	IEEE Transaction Journal draft	Not yet	Not yet	-
12.	Confirmation document	Not yet	Not yet	-