2.

Results Table:

System	Query	P @ 5	nDCG @ 5
BM25	espresso	0.6	0.7328
BM25	turkish coffee	0.6	0.7328
BM25	making a decaffeinated coffee	0.6	0.906
BM25	can I use the same coffee grounds twice	0.4	0.9197
TF-IDF	espresso	0.8	0.9047
TF-IDF	turkish coffee	0.6	0.7328
TF-IDF	making a decaffeinated coffee	0.6	0.9675
TF-IDF	can I use the same coffee grounds twice	0.6	1
VSM	espresso	0.8	1
VSM	turkish coffee	0.8	0.9558
VSM	making a decaffeinated coffee	0.2	1
VSM	can I use the same coffee grounds twice	0.6	1

When looking through the results, I noticed that the TF-IDF and BM25 models resulted with almost the exact same documents. They were ranked differently, but they still pulled the same docs. They both also produced more questions with the shorter queries but produced more answers with longer queries. This would make sense since shorter queries are more general which is also the case with questions, whereas longer queries are much more precise much like answers.

For all the systems it seems like the precision was higher with shorter queries. This is mostly the case with VSM, but still slightly the case with BM25 and TF-IDF models.

Even though the VSM model had more sporadic precision, it's ability to rank documents, represented by the nDCG values, was almost perfect. This is unlike the VSM model whose ranking score stayed high even if it was a long or short query.

It seems like for the TF-IDF and BM25 models, even though their precision fell a little bit as the queries got longer, their rankings got better.