TablePilot: Recommending Human-Preferred Tabular Data Analysis with Large Language Models

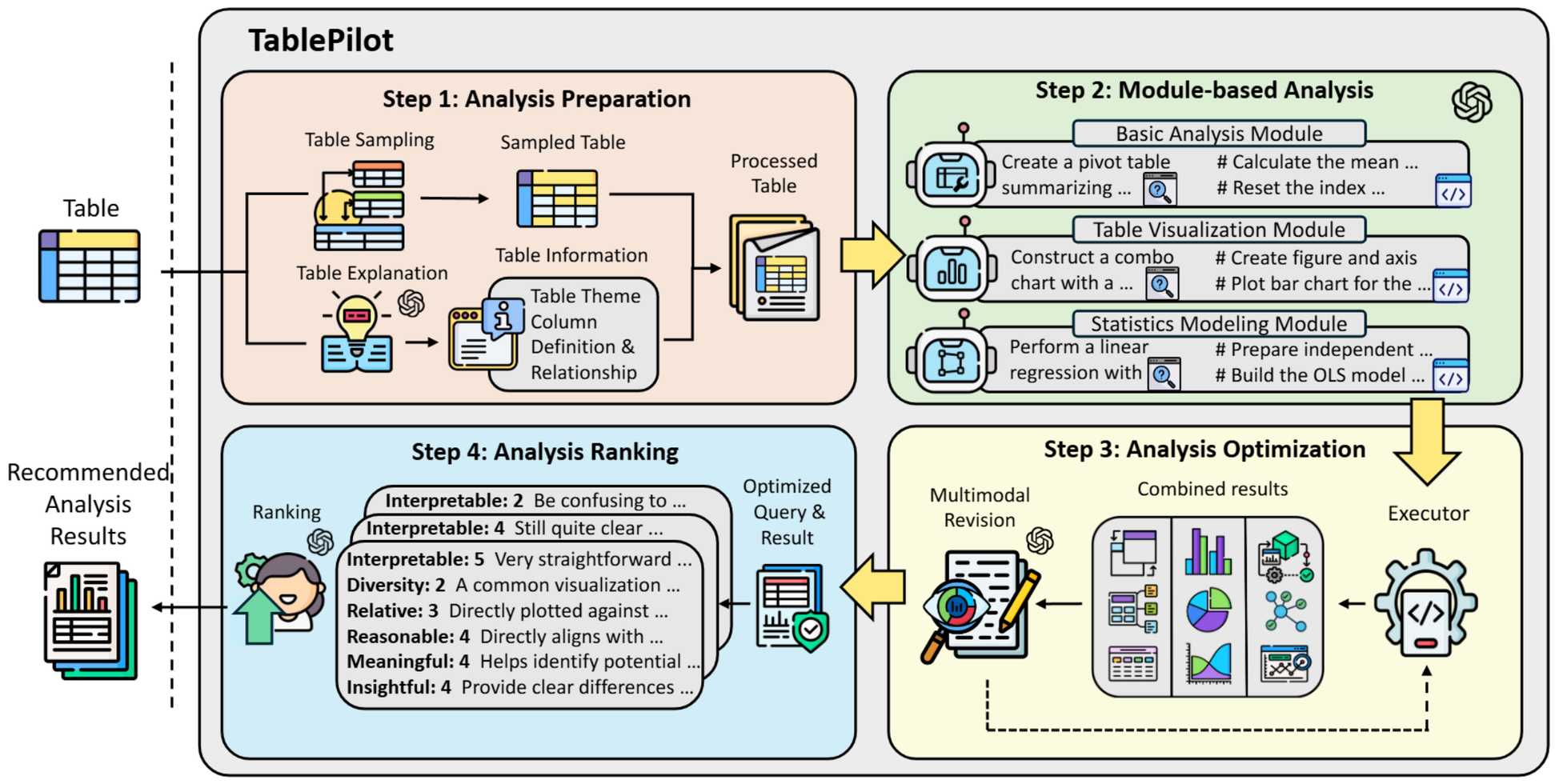
Deyin Yi, Yihao Liu, Lang Cao, Mengyu Zhou, Haoyu Dong, Shi Han, Dongmei Zhang

Microsoft

Shanghai University of Finance and Economics, Peking University, University of Illinois Urbana-Champaign, Microsoft

Abstract

Tabular data analysis is crucial in many scenarios, yet efficiently identifying relevant queries and results for new tables remains challenging due to data complexity, diverse analytical operations, and high-quality analysis requirements. To address these challenges, we aim to recommend query–code–result triplets tailored for new tables in tabular data analysis workflows. In this paper, we present TablePilot, a pioneering tabular data analysis framework leveraging large language models to autonomously generate comprehensive and superior analytical results without relying on user profiles or prior interactions. Additionally, we propose Rec-Align, a novel method to further improve recommendation quality and better align with human preferences. Experiments on DART, a dataset specifically designed for comprehensive tabular data analysis recommendation, demonstrate the effectiveness of our framework. Based on GPT-4o, the tuned TablePilot achieves 77.0% top-5 recommendation recall. Human evaluations further high-light its effectiveness in optimizing tabular data analysis workflows.



Framework

- ***Step 1:*** Sample the input table and generate corresponding explanations for its structure and content.
- ***Step 2:*** Generate query and code for modules involving basic analysis, table visualization, and statistics modeling.
- ***Step 3:*** Optimize the quality of <query, code, result> triplets.
- ***Step 4:*** Score and rank the optimized results based on multiple criteria to recommend the top-K analysis.

Main contributions

- ***Tabular Data Analysis Recommendation Framework.*** We propose TablePilot, a framework for zero-turn recommendation in tabular data analysis, encompassing a comprehensive set of analytical operations. We also contribute DART, a dataset to support and validate our framework.
- ***Table Augmentation and Multimodal Revision.*** We introduce two additional steps to enhance the accuracy of analysis results, applied before and after core analysis. These steps incorporate sampling, explanation, and multi-faceted refinement.
- ***Better Alignment with Human Data Analysis Preferences.*** We develop Rec-Align, a method designed to align recommendations with human analytical preferences, further enhancing the quality and practical utility of the recommended results.

Methodology

Our TablePilot framework processes tabular data through four key steps, formalized as $TablePilot(T) = A$.

1. ***Analysis Preparation:*** Sample a subset $Sampling(T_{a\times b}) = T'_{a'\times b'}$ and generate metadata $Explanation(T) = E$.
2. ***Module-based Analysis:*** Apply specialized modules M_K (basic analysis, visualization, statistical modeling) to produce query-code pairs $M_K(T', E) = (q_k, c_k)$.
3. ***Analysis Optimization:*** Execute code to obtain results r, then refine triplets a' using $OptimizeA$ or $OptimizeB$ based on error status.
4. ***Analysis Ranking:*** Score triplets via multi-dimensional criteria, aggregate scores s, and select top-k results $A_{k'} = Top_k(Rank(\{(q', c', r')_i\}_{i=1}^n))$.

We enhance Tablepilot using two key techniques:
Analysis SFT: Trains M_{ba}, M_{dv}, M_{sm} to generate accurate query-code pairs. ***Rank SFT:*** Improves the ranking module's adherence to evaluation criteria. ***Rank DPO :*** Aligns ranking scores with human analytical preferences using preference data.

Case Study

Airport Code	Year	Month	Domestic Passengers	International Passengers	Total Passengers
ACY	2015	Jan	98177	90	98267
ACY	2015	Feb	96431	65	96496
			.....		
JFK	2015	Mar	2209055	2208031	4417086
			.....		
JFK	1977	Jun	969248	937148	1906396
JFK	1977	Jul	1171658	1201676	2373334
			.....		

Deep Analysis based on LLM

```
Table Theme: Annual and Monthly Passenger Traffic Statistics for Major U.S. Airports
Generating various analysis and code ...
Correcting the analysis results...
Evaluating, returning a ranked list...
```

Analysis Type: SQL - Like

Group by 'Airport Code', calculate the average domestic passengers, and sort in descending order

Airport Code	Domestic Passengers
JFK	1388556
ACY	102394.4
...	

Show Code

Analysis Type: Visualization

Create a line chart to analyze the trend of total passengers over the years for each airport

Show Code

Analysis Type: Modeling

Build a statistical model to forecast the total passengers at JFK airport for the next 12 months

Date	Forecasted Passengers
2016 - 01	4276508
2016 - 02	3853416
...	

Show Code

Experiments

Method	Basic Analysis			Data Visualization			Statistics Modeling			Overall		
	R@3	R@5	R@N	R@3	R@5	R@N	R@3	R@5	R@N	R@3	R@5	R@N
GPT-4o												
Baseline	13.00	20.11	42.00	17.57	26.30	53.40	15.08	27.08	56.67	38.11	52.11	80.00
Vanilla	14.05	21.07	50.67	35.84	48.81	69.37	15.48	38.91	59.58	53.51	70.90	87.67
Analysis SFT + Rank Vanilla	15.67	22.33	55.33	43.88	53.06	70.41	20.00	30.42	61.25	59.00	72.67	89.00
Analysis SFT + Rank SFT	15.67	28.00	55.33	41.84	53.06	70.41	21.25	38.33	61.25	58.00	74.33	89.00
Analysis SFT + Rank SFT-V	15.33	25.67	55.33	**44.22**	54.42	70.41	16.25	45.83	61.25	61.00	75.00	89.00
Analysis SFT + Rank SFT & DPO	**19.33**	**30.00**	55.33	42.86	52.72	70.41	20.42	42.08	61.25	61.33	76.00	89.00
Analysis SFT + Rank SFT-V & DPO	17.67	26.00	55.33	43.88	**54.78**	70.41	**22.92**	**47.08**	61.25	**63.00**	**77.00**	89.00
GPT-4o-mini												
Baseline	15.99	24.94	35.33	27.33	39.33	44.22	3.61	6.67	35.33	29.33	42.44	62.67
Vanilla	8.67	10.67	38.33	40.48	50.34	56.12	5.54	10.83	38.33	45.33	56.67	78.33
Analysis SFT + Rank Vanilla	13.00	**57.14**	46.67	44.22	25.33	64.29	1.67	10.42	59.58	52.00	68.67	85.00
Analysis SFT + Rank SFT	**24.91**	34.33	46.67	34.15	45.24	64.29	12.02	32.08	59.58	56.66	71.67	85.00
Analysis SFT + Rank SFT-V	16.00	24.33	46.67	**46.60**	**54.08**	64.29	22.50	43.33	59.58	61.00	75.00	85.00
Analysis SFT + Rank SFT & DPO	21.33	32.67	46.67	42.86	50.34	64.29	16.25	27.05	59.58	60.33	73.67	85.00
Analysis SFT + Rank SFT-V & DPO	21.00	29.00	46.67	40.14	51.02	64.29	**22.92**	**49.17**	58.58	**62.33**	**76.67**	85.00
Phi-3.5-vision												
Baseline	3.00	4.00	5.00	1.36	3.40	4.08	0.00	0.00	0.42	4.33	7.00	8.67
Vanilla	1.43	1.79	13.33	1.83	1.83	3.74	3.12	3.12	7.92	5.73	6.09	21.67
Analysis SFT + Rank Vanilla	3.77	3.77	24.00	**3.83**	**4.53**	9.52	18.45	19.31	32.50	20.89	21.58	47.67
Analysis SFT + Rank SFT	**6.85**	14.04	24.00	2.79	4.18	9.52	15.88	22.75	32.50	20.89	32.19	47.67
Analysis SFT + Rank SFT-V	5.14	13.01	24.00	1.74	3.14	9.52	**19.31**	21.89	32.50	21.23	30.14	47.67
Analysis SFT + Rank SFT & DPO	**8.90**	**15.07**	24.00	1.74	3.83	9.52	18.88	23.61	32.50	**23.97**	**32.88**	47.67
Analysis SFT + Rank SFT-V & DPO	7.53	14.38	24.00	1.74	2.09	9.52	**19.31**	**25.32**	32.50	23.63	32.19	47.67

ACL 2025 VIENNA