

# 実験結果報告書 (Thesis Report Draft)

## 1. 定量評価 (Quantitative Evaluation)

本節では、ログデータおよびアンケート回答に基づく定量的な分析結果を述べる。

### 1.1 作業効率と探索行動への影響 (Efficiency and Exploration)

推薦機能の有無 (Rec vs. No Rec) およびタスクの種類 (UFO vs. Wine) による、平均作業時間、平均作成ビュー数、および平均ユニークビュー数 (探索の多様性) の比較結果を Table 1 に示す。

Table 1: Performance metrics by task and condition.

Task	Condition	\$n\$	Time (min) [Mean]	Created Views [Mean]	Unique View Types [Mean]	Rec Usage (%)
UFO	Rec	6	22.85	6.2	4.50	68.6%
	No Rec	6	20.78	6.8	4.00	-
Wine	Rec	6	24.51	6.8	4.33	46.2%
	No Rec	6	26.49	7.2	3.67	-

#### 作業時間に関する考察

Table 1 に示されるように、**Wine タスク**においては推薦あり (Rec) 条件で平均作業時間が約 2 分短縮された (26.49分  $\rightarrow$  24.51分)。一方、**UFO タスク**においては逆に約 2 分の増加が見られた (20.78分  $\rightarrow$  22.85分)。この差異は、タスクおよび対象データの複雑性に起因すると考えられる。Wine データセットはタスクで分析対象となるデータ (品種、国など) が多く、適切な可視化手法の選択が困難であるため、推薦による支援が効率化に寄与したと推測される。対照的に、UFO データセットは時系列分析が主となるため可視化の定石が明確であり、推薦の確認作業がオーバーヘッドとなった可能性がある。

#### 探索の多様性に関する考察

作成されたチャートタイプの種類数 (Unique View Types) に着目すると、両タスクにおいて Rec 条件で値が増加している (UFO: +0.50, Wine: +0.66)。これは、推薦機能がユーザに対し、自発的には選択しにくい多様な可視化手法の利用を促進したことを示唆している。

詳細なチャートタイプの利用内訳を Table 2 に示す。**Wine タスク**では、No Rec 条件で支配的であった「棒グラフ (18)」や「数値 (12)」の利用が Rec 条件で減少し (棒: 10, 数値: 8)、代わりに **Gauge (4)** や **Scatter (3)** といった高度なチャートが出現した。**UFO タスク**においても、No Rec 条件では見られなかった **Pivot (5)** が Rec 条件で利用されており、推薦機能が分析の視点 (Viewの多様性) を広げる効果を持つことが定量的に確認された。

Table 2: Comparison of created chart types and recommendation counts.

Chart Type	UFO (No Rec)	UFO (Rec: Used)	UFO (Rec: Shown)	Wine (No Rec)	Wine (Rec: Used)	Wine (Rec: Shown)
------------	--------------	-----------------	------------------	---------------	------------------	-------------------

Chart Type	UFO (No Rec)	UFO (Rec: Used)	UFO (Rec: Shown)	Wine (No Rec)	Wine (Rec: Used)	Wine (Rec: Shown)
Bar (棒グラフ)	10	4	17	18	10	0
Pie (円グラフ)	9	12	38	6	9	41
Line (折れ線)	8	7	6	-	-	31
Scalar (数値)	6	-	0	12	8	0
Map (地図)	6	6	0	-	-	4
Pivot (ピボット)	-	5	25	5	6	18
Gauge (ゲージ)	-	-	27	-	4	27
Scatter (散布図)	-	-	13	-	3	35

2. 定性評価 (Qualitative Evaluation)

アンケートの自由記述回答および定量的回答に基づき、推薦機能がユーザの行動変容に与えた影響を分析した。

2.1 ユーザビリティと主観的評価（アンケート定量結果） (Usability and Subjective Evaluation)

システム全体のユーザビリティ（SUS）の各質問項目のスコアを Table 3 に、タスクごとの成果物評価および推薦機能への評価を Table 4 に示す。

Table 3: System Usability Scale (SUS) item scores and total statistics (\$n=12\$).

#	Question Item (System Usability Scale)	Score [Mean]
Q1	このシステムを頻繁に使いたいと思うか	4.25
Q2	このシステムは不必要に複雑だと感じたか	2.00
Q3	このシステムは使いやすいと思ったか	3.75
Q4	このシステムを使うには技術者のサポートが必要だと思うか	3.50
Q5	このシステムの様々な機能はうまく統合されていると感じたか	4.33
Q6	このシステムには矛盾が多すぎると感じたか	1.25
Q7	ほとんどの人がこのシステムをすぐに使いこなせるようになると思うか	2.83
Q8	このシステムは非常に扱いにくいと感じたか	1.75

#	Question Item (System Usability Scale)	Score [Mean]
Q9	このシステムを使うことに自信を感じたか	3.75
Q10	このシステムを使い始める前に、多くのことを学ぶ必要があると感じたか	2.67
Total	Overall SUS Score	69.4

SUS スコアの算出にあたっては、Brooke [1] の定義に従い、奇数項目（ポジティブな質問）は「回答値 - 1」、偶数項目（ネガティブな質問）は「5 - 回答値」としてスコア化し、その合計を 2.5 倍することで 0-100 のスケールに換算している。本実験における平均スコアは 69.4 であり、Sauro & Lewis [2] が提唱する一般的な平均基準（68）を上回る結果となった。特に Q6（矛盾）や Q8（扱いづらさ）の生スコアが低いことは、システムへの否定的な評価が少なかったことを示しており、これらが寄与スコアとして総合得点を押し上げている。

**References** [1] Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7. [2] Sauro, J., & Lewis, J. R. (2016). Quantifying the User Experience: Practical Statistics for User Research (2nd ed.). Morgan Kaufmann.

Table 4: Subjective scores for tasks and recommendation features (1-5 scale).

Category	Question Item	Rec [Mean]	No Rec [Mean]
Output Quality	作成した Dashboard は、クライアントにとって情報を読み取りやすいものになったと思うか	4.08	3.67
	作成した Dashboard は、タスクの要件を達成していると思うか	4.33	4.42
Recommendation	推薦システムから得られた推薦は、Dashboard 作成の役に立ったと思うか	3.92	-
	推薦システムから得られた推薦は、意外性のあるものだったと思うか	3.58	-

Table 4 より、推薦機能あり（Rec）の条件では、作成されたダッシュボードの「読み取りやすさ」に対する自己評価が向上する傾向が見られた（No Rec: 3.67 \$to\$ Rec: 4.08）。また、推薦自体の有用性も 3.92 と肯定的に評価されている。

2.2 推薦機能によるポジティブな効果

推薦機能が奏功した要因として、Table 5 に示すように「選択の効率化」と「新規視点の受容」が挙げられる。

Table 5: Factors contributing to positive impact with individual metrics (Rec vs. No Rec).

Factor	User	Rec Rate	Task (Rec / No Rec)	Time (Rec / No Rec)	Views (Rec / No Rec)	Typical User Comment (Excerpt)
選択の効率化	つばさ	80%	UFO / Wine	16.1 / 25.9	5 / 7	「数あるグラフの中から推薦によって限定してくれていることで、グラフを選択しやすかった」

Factor	User	Rec Rate	Task (Rec / No Rec)	Time (Rec / No Rec)	Views (Rec / No Rec)	Typical User Comment (Excerpt)
初心者への支援	ゆうご	83%	UFO / Wine	22.9 / 25.6	6 / 7	「初心者にとっては効率的...一方で、見慣れない図表...使用方法も併せて推薦してほしい」
視点の受容	たくみ	83%	Wine / UFO	23.5 / 22.6	6 / 7	「ゲージや散布図は当初使う予定はなかったが推薦を受け入れた」

2.3 推薦機能によるネガティブな効果

一方、作業時間の増加や迷いを生んだ要因として、Table 6 に示す「解釈の負荷」や「システムへの疑義」が確認された。

Table 6: Factors contributing to negative impact with individual metrics (Rec vs. No Rec).

Factor	User	Rec Rate	Task (Rec / No Rec)	Time (Rec / No Rec)	Views (Rec / No Rec)	Typical User Comment (Excerpt)
解釈の負荷	いたい	71%	UFO / Wine	24.3 / 20.3	7 / 8	「仮説・説明とともにダッシュボードが提示されると、人間側の仮説思考の負担が減るので楽だと思った」
推薦への疑義	たなか	25%	Wine / UFO	25.5 / 16.0	8 / 6	「折れ線グラフをこのデータで使うの難しくないか」
探索の深化	はなわ	83%	Wine / UFO	31.7 / 20.7	6 / 8	「推薦の機能を参考にしたが、追加分析では...自身でグラフの選択を行った」

3. 結論 (Conclusion)

以上の結果より、本システムは特に複雑なデータセット（Wine）における作業効率の向上と、分析視点の多様化に寄与することが示された。一方で、推薦意図の解釈に伴うコストが課題として確認されたため、今後は推薦理由の説明（Explanation）機能の拡充が求められる。