

高度なデータ分析(GPT-4o, Code interpreter) 操作マニュアル

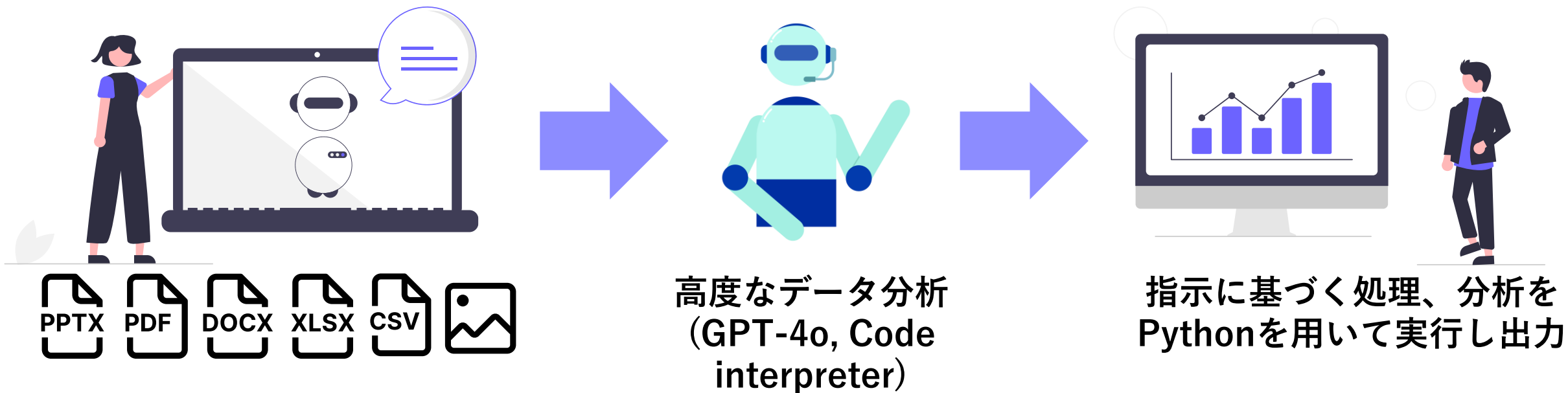
目次

- 01. 高度なデータ分析(GPT-4o, Code interpreter)とは P.4.
- 02. 基本的な操作方法 P.8
- 03. 活用例 P.13

高度なデータ分析(GPT-4o, Code interpreter) とは

高度なデータ分析(GPT-4o, Code interpreter) とは

自然言語の指示文でプログラミング（Python）によるファイル処理、データ分析、機械学習等ができる機能



プログラミングの知識がなくても、プログラミングを用いた複雑な処理や機械学習による分析などをexaBase 生成AIのUI上で実行できます

ファイル添付したデータ（指示文・プロンプト内への記述でもOK）を下記のように処理させ、GPT-4oと掛け合わせて出力をさせることができます

活用方法	内容	具体的なユースケース例
データ分析	CSVやExcelファイルを読み取り、集計やグラフ作成ができます。	売上データを月別に集計し、グラフで表示する。 アンケート結果のクロス集計による相関分析を行い、散布図を作成する。
データ整形	CSVやExcelファイルのデータの加工ができます。	欠損値や表記ゆれ、外れ値のあるデータを修正する。 平均、標準偏差などの集計データ追加させ出力させる。 異なるシートのデータを1つの表にまとめる。
テキスト処理	テキストデータから必要な情報を抽出したり、形式を整えることができます。	お客様アンケートのフリーテキストから「ポジティブ」「ネガティブ」分類し比率を試算したり、頻出単語を抽出し傾向を分析させる。
自動計算	数学的な計算や回帰分析などの統計処理ができます。	過去の売上データから季節性やトレンドを加味したモデルを構築し売上予測をさせる。
ファイル変換	word→パワーポイント、csv→wordなど形式の変換ができます	wordで書いた台本をパワーポイントでプレゼン資料に変換させる。

csv、xlsx、pptx、txt、xml、json、html等の多彩なファイルで入出力が可能です

CSVファイルを添付

添付ファイルはCSV UTF-8の形式です。このファイルデータのうちサーバに関連するトラブルを抽出し、その発生原因のランキングベスト5を出力してください

コンピューティング研究所トラブル_sample0325
ファイル

```
import pandas as pd

# Load the CSV file to check its contents
file_path = '/mnt/data/file-HLwx8TAoanCzMj77oT7Eug'
data = pd.read_csv(file_path)

# Display the first few rows and columns of the data for overview
data.head(), data.columns
```

データには以下のような列が含まれています。

- 発生日付: トラブルが発生した日付
- 発生部署: トラブルが発生した部署名
- トラブル概要: トラブルの内容の概要
- 発生原因: トラブルの原因
- 対策: トラブルを解決するための対策
- 損失金額 (千円): トラブルによる損失金額
- 原因分類: トラブルの原因を分類したカテゴリ

次に、このデータから「サーバ」に関連するトラブルを抽出し、その中で「発生原因」を集計してランキングトップ5を抽出します

集計させたファイルをCSVファイルで出力 ※

当該データのB列 発生部署ごとの原因分類の件数を集計してCSVファイルで出力してください

```
# 発生部署ごとに原因分類の件数を集計
cause_by_department = data.groupby(['発生部署', '原因分類']).size().unstack(fill_value=0)

# CSVファイルとして出力
output_file_path = '/mnt/data/発生部署ごとの原因分類集計.csv'
cause_by_department.to_csv(output_file_path)

output_file_path
```



「発生部署ごとの原因分類の件数」を集計してCSVファイルに出力しました。以下のリンクからダウンロードできます。

[発生部署ごとの原因分類集計.csv](#)

tmpsxv6jzbz.csv
ファイル



発生部署	その他	システム障害	ヒューマンエラー	外部要因	自然災害	設計不備	運用ミス
人事部	0	0	0	0	5	0	0
企画部	1	1	5	2	2	1	1
品質管理部	2	2	2	1	2	0	2
営業部	1	0	2	0	0	2	1
情報システム部	0	2	1	1	2	0	0
生産部	1	0	1	3	1	1	1
研究部	1	1	1	0	2	1	0
経理部	1	2	0	3	1	1	0
総務部	1	3	2	2	0	1	1
製造管理部	1	2	3	1	1	1	3
販売促進部	3	1	0	2	0	1	2
開発部	0	1	1	2	1	0	1

データを可視化させることが可能です

相関の可視化を依頼

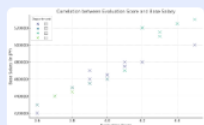
評価スコアと給与の相関を視覚化してください

こちらが、評価スコアと基本給の関係を示す散布図です。データポイントは部署ごとに色分けされており、評価スコアが高い人ほど基本給が高い傾向が見られる可能性があります。

観察点:

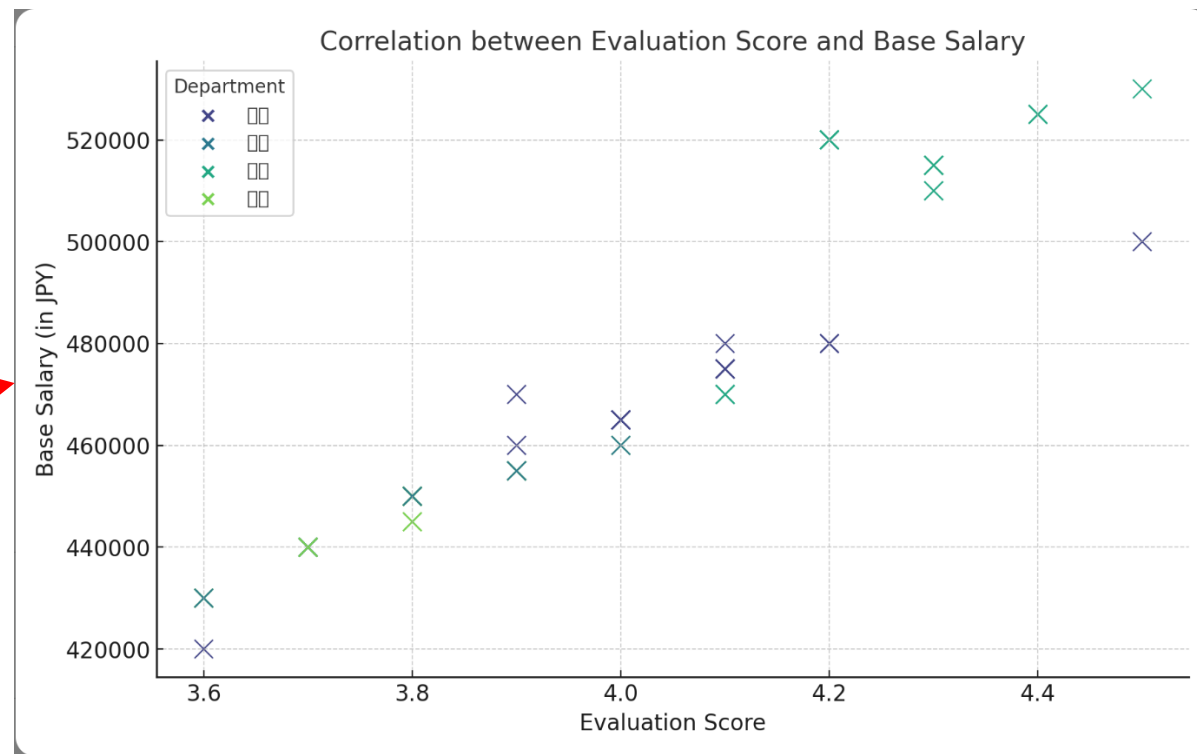
- 部署間で基本給の分布に違いがあります。
- 全体的に評価スコアが高くなるに連れ基本給も高くなるように見えますが、いくつか例外もあります。

その他このデータ分析で進めてほしい方向性がありますか？



視覚化、棒グラフ、円グラフの作成などを依頼

グラフ形式で可視化



相関の視覚化を依頼したため、散布図で出力された

基本的な操作方法

「高度なデータ分析(GPT-4o, Code interpreter)」を選択し、ファイルを添付※

※先に指示文・プロンプトを入力しても問題ありません



左上のプルダウン選択から「高度なデータ分析 (GPT-4o, Code interpreter)」を選択

指示文・プロンプト入力欄の左のアイコンを選択し
分析したいファイルを選択
※ドラッグ&ドロップで添付も可能

指示文・プロンプト入力の際はファイル形式またはファイル名を指定※して会話を送信

※ファイル名指定については次頁に詳細を記載

高度なデータ分析(GPT-4o)

お疲れ様です!
どんな作業をお手伝いしましょうか?

校正・翻訳・要約
校正・チェックする

校正・翻訳・要約
翻訳する

文字起こし
(音声/画像)
通話記録の文字起こし・要約・課題抽出 ※音声ファイル添付必須

校正・翻訳・要約
要約する

VBA・Excel・データ分析
Excel関数の作成

テンプレートを使用

改行するには shift + Enter を押してください

ID, 名前, 年齢, 部署, ...
ドキュメント

ファイル名 ID, 名前, 年齢, 部署, 評価スコア, 基本給, ボーナス.csv を読み込み、多角的に分析を行ってください。

ファイルが添付されていることを確認し、ファイル名を指定して指示文・プロンプトを入力。右の「会話を送信」アイコンをクリック

部署ごとの統計情報を分析しました。

部署別の特徴：

1. 営業
 - 年齢: 平均 33.0歳, 最年少 26歳, 最年長 42歳
 - 評価スコア: 平均 4.0, 最大値 4.5
 - 基本給: 平均 469,615円 (部署内で最小～最大は 420,000～500,000円)
2. 経理
 - 年齢: 平均 35.6歳 (他の部署より高め)
 - 評価スコア: 平均 4.2 (全体的に高評価)
 - 基本給: 平均 508,333円 (部署での最大給与 530,000円)
3. 総務
 - 年齢: 平均 33.5歳
 - 評価スコア: 平均 4.0
 - 基本給: 平均 448,542円 (部署内での幅が比較的狭い)
4. 開発
 - 年齢: 平均 30.8歳 (最も若い傾向)
 - 評価スコア: 平均 3.8 (他部署より若干低め)
 - 基本給: 平均 446,346円

次のステップ:

- 評価スコア と 基本給 の散布図で関係性を可視化しますか？
- その他に見たい視点や分析ポイントがあれば教えてください。

分析結果が出力。最後に出力されているフォロー提案を参考に可視化、深掘り、異なる分析をリクエスト可能

データを読み込ませるコツは、指示文・プロンプトにファイル名を指定

ファイル形式を指定せず読み込み

添付のファイルの読み込みをお願いいたします

com_incident_sample0325.csv
ファイル

```
# Let's first try to identify what kind of file has been uploaded by the user.  
  
file_path = '/mnt/data/file-SojFgLj6jvDgLjdnFJMwJJ'  
  
import mimetypes  
  
file_type, _ = mimetypes.guess_type(file_path)  
file_type
```

お客様にアップロードいただいたファイルの種類を解析できませんでした。このファイルの具体的な形式（例: Excel, CSV, テキストなど）をお知らせいただけますか？または、簡単に解釈できるようにファイルの内容について教えていただけますと幸いです。

読み込みエラーが発生することがあります

ファイル名を指定して読み込み

添付したファイル名com_incident_sample0325.csvを読み込んでください

com_incident_sample0325.csv
ファイル

```
import pandas as pd  
  
# ファイルパス  
file_path = '/mnt/data/file-FbxZdhkWC7uq3BA5uanYM1'  
  
# CSVファイルを読み込む  
data = pd.read_csv(file_path)  
  
# データの先頭5行を確認  
data.head()
```

ファイル「com_incident_sample0325.csv」を正しく読み込みました。以下はデータの先頭5行です。

発生日付	部署	カテゴリー	事象	発生原因	対策	損失金額（千円）
2024/3/10	情報システム部	外部要因	アプリ強制終了頻発。データ破損が見つかるが原因で外部から攻撃	外部攻撃の防御不足のため過電流で基板破損	冗長構成を強化する、外部ベンダと協業する	1970

スムーズに読み込み成功

※データに欠損値があるなど問題があると、時間がかかるケースもあります

さらに添付するファイル名を半角英数字、データの内容を指示文・プロンプト内に入力すると分析の精度が良くなる傾向があります

半角英数字のファイル名、データの内容を入力し 多面的な分析を依頼

添付のファイル名com_incident_sample0325.csvは、コンピューティング関連の研究所におけるインシデントの記録データです。このデータを読み込み多面的な分析を行ってください

com_incident_sample0325.csv
ファイル

出力結果

カテゴリ別の分析結果は以下の通りです：

1. インシデント件数:

- 「ヒューマンエラー」が最多の18件記録され、合計損失額は約17,808千円です。
- 「外部要因」と「自然災害」がそれぞれ17件で続き、それらの合計損失額は約15,249千円と15,242千円ではほぼ同程度です。

2. 平均損失額が高いカテゴリ:

- 「設計不備」の平均損失額が最も高く、1,326千円を記録しています。
- 次いで「運用ミス」が1,053千円、「その他」も約933千円です。

「ヒューマンエラー」と「設計不備」が特に大きな問題となっている可能性が見えます。

次に、時間的な分析として、発生日付ごとのインシデント傾向を確認しましょう。頻発している時期や季節性パターンがないか検討します。



部署ごとの分析結果は以下の通りです：

1. インシデント件数が多い部署:

- 「企画部」は最多の13件で、損失額の合計が約13,513千円と大きな影響を受けています。
- 「製造管理部」が次いで12件のインシデントが発生しており、損失額の合計も12,320千円と高水準です。

2. 平均損失額が高い部署:

- 損失額の平均が最も高いのは「情報システム部」で、1,370千円を超えています。
- 「営業部」や「生産部」も約1,200千円という高めの平均損失額を記録しています。

次に、カテゴリ別でのインシデント数および損失額を分析し、さらに詳細な関連性を明らかにします。

損失額が高額な上位10件の分析結果は次の通りです：

1. 損失額が最大の事案:

- 2024年3月10日に「情報システム部」で発生した事案では、損失額が1,970千円です。
- 原因は「外部攻撃の防御不足」、対策として「冗長構成の強化、外部ベンダとの協業」を挙げています。

2. 他の高額損失事案:

- 2024年2月11日（企画部、1967千円）：設定ミスによるライブラリ更新の問題。
- 2024年12月19日（情報システム部、1930千円）：ライブラリ更新不備や機器の異常加熱。
- 2024年4月1日（研究部、1890千円）：コンフィグ衝突やヒューマンエラーが原因。

3. 共通する原因と対策:

- 多くのケースで「ライブラリ更新不備」「外部攻撃の防御不足」「ヒューマンエラー」などが見られる。
- 対策として「冗長構成の強化」「監視体制の拡充」「定期点検の徹底」などが頻出しています。

これらの上位事案から、特に注力すべき領域は「セキュリティ強化」「定期的な更新と検査」「操作ミス防止の運用手順改善」であると言えます。

他に特定したい視点や分析はありますか？例えば、特定カテゴリに絞った分析や別の統計視点での要約なども可能です。

活用例

※いくつかのユースケースは標準テンプレートでご提供予定です

欠損値や表記ゆれがあるデータを分析できる状態にする

欠損値や表記ゆれのあるのデータ

名前	ID	生年月日	身長(cm)	体重(kg)	血糖値(mg/dL)	収縮期血圧(mmHg)	拡張期血圧(mmHg)
渡辺翔太	8974-6499	2001年3月18日		71.1		120.52	
鈴木美咲	9465-8268	1987/6/17			78.4780	97.88	105.25
渡辺太郎	2325-2349	1976/11/24				172.49	64.31
佐藤 翔太	3089-1689	1964/8/4	163.5		147.6379		
佐藤一郎	4549-1894	1999/2/1				143.67	74.86
渡辺花子	39830859	1989/8/16	162.3206			102.75	
田中 一郎	7382-4800	1967/2/28		64.31	86.84		
高橋美咲	37168934	19700403	178.4			99.3	65.6
田中 太郎	93113945	1989/2/17			83.8	92.5	76.8
田中 翔太	7749-9187	1993/7/26				127.1	
鈴木 翔太	69091497		159.7	51.2			97.7
鈴木一郎	52790096	1998/9/9			133.7	177.8	
田中花子	5888-2030		161.0		84.4		116.7
佐藤 花子	1656-2434			94.4	136.2		105.8
鈴木 美咲	3582-6291	1989/11/19			103.1379719		
鈴木 一郎	5574-4613	1973/11/18	174.9		123.2		
佐藤翔太	50418366	1993/9/25	165.8				99.3
高橋 太郎	7310-7148	1967/10/1	157.9	71.4	86.2		89.9
田中美咲	2909-4055		164.3		104.3		
鈴木 太郎	2631-2206			71.0			94.4
田中 美咲	5769-3477	1990/10/28	159.9	78.7		153.5	118.5
田中 太郎	14451472		168.5886806			124.3889091	99.0589336

データクレンジングを依頼



あなたは非常に優秀なデータサイエンティストです。以下の内容をコードインタープリターの機能を使って実行してください。

1. ファイルの取り込み

- 解析対象のファイル（CSV、Excel など）をアップロードしてください。
 - ファイル名： data前処理デモsample_0317
 - 未指定の場合は、アップロードを確認した上で適切にファイルを読み込みます。
- もしファイルが未添付の場合は、その旨を教えてください。添付ファイルの有無とファイル名を確認後、解析を進めます。

2. 目的

- データを前処理・クレンジングし、不要なデータや欠損値を適切に処理して整形し、ダウンロード可能な形式（CSV または Excel）にしたい。

3. データの概要把握

- ファイルを読み込んだ後、以下のサマリーを出力してください：
 - 行数・列数
 - 各列のデータ型
 - 欠損値の分布
 - 重複行の有無（あれば重複行数）
 - 数値列の統計量（平均・標準偏差・最小値・最大値など）

4. 前処理・クレンジング手順

4-1. 不要な列・重複データの削除

- 指示例：削除したい列 = ["example_column"]
- 未指定の場合は、列の削除や重複データの削除は行いません。

修正後のデータをCSVで出力

クレンジング後のデータをCSVで出力

```
# 1. 名前列の整形: 苗字と名前の間を半角スペースに統一
data['名前'] = data['名前'].str.replace(r'[\s ]+', ' ', regex=True)

# 2. 生年月日列をdatetime型に変換、欠損値は最頻値で補完
data['生年月日'] = pd.to_datetime(data['生年月日'], errors='coerce')
most_frequent_birth_date = data['生年月日'].mode().iloc[0] # 最頻値
data['生年月日'].fillna(most_frequent_birth_date, inplace=True)

# 3. 数値列の欠損値を平均値で補完
numeric_columns = data.select_dtypes(include=['float64']).columns
for col in numeric_columns:
    mean_value = data[col].mean()
    data[col].fillna(mean_value, inplace=True)

# 修正後の概要再確認
cleaned_data_info = {
    "shape": data.shape,
    "missing_values": data.isnull().sum().to_dict(), # 欠損値数
```

整形後のデータをCSV形式で保存しました。以下のリンクからダウンロード可能です。

[整形後データをダウンロード](#)

次に進む処理やさらに分析を希望する場合は、指示をお知らせください！

tmp7si_ex94.csv
ファイル



データクレンジング後のデータ

名前	ID	生年月日	身長(cm)	体重(kg)	血糖値(mg/dL)	収縮期血圧	拡張期血圧
渡辺翔太	8974-6499	1991/12/25	169.4	71.1	111.1	121	86
鈴木美咲	9465-8268	1987/6/17	169.4	76.3	78.5	98	105
渡辺太郎	2325-2349	1976/11/24	169.4	76.3	111.1	172	64
佐藤 翔太	3089-1689	1964/8/4	163.5	76.3	147.6	135	86
佐藤 一郎	4549-1894	1999/2/21	169.4	76.3	111.1	144	75
渡辺花子	39830859	1989/8/16	162.3	76.3	111.1	103	86
田中 一郎	7382-4800	1967/2/28	169.4	64.3	86.8	135	86
高橋美咲	37168934	1991/12/25	178.4	76.3	111.1	99	66
田中 太郎	93113945	1989/2/17	169.4	76.3	83.8	92	77
田中 翔太	7749-9187	1993/7/26	169.4	76.3	111.1	127	86
鈴木 翔太	6909-1497	1991/12/25	159.7	51.2	111.1	135	98
鈴木 一郎	5279-0096	1998/9/9	169.4	76.3	133.7	178	86
田中花子	5888-2030	1991/12/25	161.0	76.3	84.4	135	117
佐藤 花子	1656-2434	1991/12/25	169.4	94.4	136.2	135	106
鈴木 美咲	3582-6291	1989/11/19	169.4	76.3	103.1	135	86
鈴木 一郎	5574-4613	1973/11/18	174.9	76.3	123.2	135	86
佐藤翔太	5041-8366	1993/9/25	165.8	76.3	111.1	135	99
高橋 太郎	7310-7148	1967/10/1	157.9	71.4	86.2	135	90
田中美咲	2909-4055	1991/12/25	164.3	76.3	104.3	135	86

修正漏れがあるケースもあるため、チェックしさらに修正依頼することで精度を高める

wordデータ等をPPTへ変換し出力

textデータを準備しPPT化を依頼

発表用台本

[発表用台本]

イントロダクション↓

皆さん、本日はお集まりいただきありがとうございます。本日は、株式会社 Exa Enterprise AI、通称 EAI が提供する革新的なソリューションと、その AI プラットフォーム「exaBase」についてご紹介させていただきます。

スライド 1: 会社概要↓

まずは、EAI の会社概要についてお話しします。同社は、クライアントのニーズに合わせた先進的な技術提案を得意としており、その実績は ISO/IEC 27001、27017、27701 など、数々の認証を取得している点で際立っています。これらは、機密情報の保護とグローバル基準への準拠を示すものです。

スライド 2: 成長と実績↓

次に、Exa グループの成長について見てみましょう。同社はこれまで 2,000 社を超える企業と成功に満ちたコラボレーションを行ってきました。この目覚ましい成長の軌跡は、彼らの効果的な戦略と、提供する価値の高さを物語っています。

スライド 3: 包括的な AI サポート↓

EAI は、AI およびデジタルトランスフォーメーションに関わるプロジェクトを、企画から設計、開発、実装まで、一貫して支援します。業界ごとに特化したチームが専門知識を活かし、「業務改善」「人材育成」「コンサルティング」など、さまざまなニーズに応えています。

PPTで出力

以下のリンクからプレゼンテーションファイルをダウンロードできます：

[PowerPointプレゼンテーションをダウンロード](#)

追加の修正やスライドデザインの要望があれば、いつでもお知らせください！



tmpeuf0d_6p.pptx
ファイル



株式会社Exa Enterprise AI 会社 紹介

革新的ソリューションとAIプラットフォーム 'exaBase'

会社概要

- 先進的な技術提案を得意とする
- ISO/IEC 27001, 27017, 27701 認証取得
- 情報保護とグローバル基準準拠

成長と実績

- 2,000社以上との成功したコラボレーション
- 効果的戦略と提供価値の高さ
- 目覚ましい成長の軌跡

包括的なAIサポート

- 企画、設計、開発、実装まで一貫支援
- 専門知識を活かす業界別チーム
- ニーズ対応: 業務改善、人材育成、コンサルティング

生成AIの活用支援

- ChatGPTの導入とカスタムツール提供
- セキュリティ重視の生成AIサービス
- サービス事業化の支援と可能性最大化

まとめ・お問い合わせ

- Exa Enterprise AIは、技術、顧客重視、未来思考の融合
- プロセス革新、新ソリューション開発、デジタルトランスフォーメーション推進
- お問い合わせ先: info@exaenterprise.ai

投稿日時	顧客ID	業種	会社規模	役職	男女	年代	コメント		
2023/1/1	C001	小売業	100名未満	役員・経営	男性	30代	初期費用が高く、コスト面で負担を感じます。		
2023/1/2	C002	不動産業	100名未満	役員・経営	男性	20代	トラブル対応のサポートが遅くて業務に支障が出ました。		
2023/1/3	C003	不動産業	100名未満	役員・経営	男性	50代	コストパフォーマンスに優れ、業務効率化に繋がりました。		
2023/1/4	C004	製造業	100名以上	役員・経営	男性	40代	期待したほど効率が上がらず、導入効果が薄いです。		
2023/1/5	C005	小売業	100名未満	管理職	男性	40代	操作が直感的で業務効率が大幅にアップしました。		
2023/1/6	C006	サービス業	100名未満	管理職	男性	40代	UIが使いやすく担当者が短時間で習得できました。		
2023/1/7	C007	公的機関	100名以上	役員・経営	男性	30代	サポートが丁寧でとても助かります。		
2023/1/8	C008	金融業	100名未満	管理職	男性	40代	期待したほど効率が上がらず、導入効果が薄いです。		
2023/1/9	C009	金融業	300名以上	役員・経営	女性	20代	機能が豊富でチーム連携がスムーズに進みます。		
2023/1/10	C010	不動産業	100名以上	一般職	男性	40代	導入が容易でスピーディに使用開始が可能でした。		
2023/1/11	C011	医療機関	100名未満	一般職	女性	40代	サポートが丁寧でとても助かります。		
2023/1/12	C012	金融業	300名以上	一般職	男性	50代	柔軟なカスタマイズが可能で多部門で活用しています。		
2023/1/13	C013	IT業	300名以上	一般職	男性	40代	新機能が定期的に追加され、とても便利だと感じます。		
2023/1/14	C014	公的機関	1000名以上	管理職	男性	30代	サポートが丁寧でとても助かります。		
2023/1/15	C015	製造業	100名未満	一般職	男性	50代	新機能が定期的に追加され、とても便利だと感じます。		
2023/1/16	C016	運輸業	1000名以上	役員・経営	男性	20代	サポートが丁寧でとても助かります。		
2023/1/17	C017	不動産業	300名以上	役員・経営	男性	40代	柔軟なカスタマイズが可能で多部門で活用しています。		
2023/1/18	C018	製造業	300名以上	役員・経営	女性	20代	コストパフォーマンスに優れ、業務効率化に繋がりました。		
2023/1/19	C019	不動産業	100名以上	一般職	男性	30代	操作が直感的で業務効率が大幅にアップしました。		
2023/1/20	C020	公的機関	300名以上	管理職	男性	40代	コストパフォーマンスに優れ、業務効率化に繋がりました。		
2023/1/21	C021	金融業	100名未満	一般職	女性	20代	アップデート後の不具合が多く、困っています。		
2023/1/22	C022	小売業	100名以上	一般職	男性	30代	新機能が定期的に追加され、とても便利だと感じます。		
2023/1/23	C023	金融業	100名以上	役員・経営	男性	40代	柔軟なカスタマイズが可能で多部門で活用しています。		
2023/1/24	C024	小売業	100名以上	一般職	男性	20代	導入が容易でスピーディに使用開始が可能でした。		
2023/1/25	C025	公的機関	100名以上	一般職	男性	20代	操作が直感的で業務効率が大幅にアップしました。		
2023/1/26	C026	医療機関	100名未満	役員・経営	男性	50代	機能が豊富でチーム連携がスムーズに進みます。		
2023/1/27	C027	公的機関	1000名以上	一般職	女性	50代	初期費用が高く、コスト面で負担を感じます。		
2023/1/28	C028	運輸業	300名以上	管理職	男性	20代	サポートが丁寧でとても助かります。		
2023/1/29	C029	製造業	1000名以上	役員・経営	男性	20代	導入が容易でスピーディに使用開始が可能でした。		
2023/1/30	C030	不動産業	1000名以上	一般職	男性	40代	サポートが丁寧でとても助かります。		
2023/1/31	C031	サービス業	100名未満	一般職	男性	20代	導入が容易でスピーディに使用開始が可能でした。		
2023/2/1	C032	IT業	1000名以上	管理職	男性	50代	機能が豊富でチーム連携がスムーズに進みます。		
2023/2/2	C033	運輸業	100名未満	管理職	男性	50代	柔軟なカスタマイズが可能で多部門で活用しています。		
2023/2/3	C034	運輸業	100名以上	管理職	男性	20代	操作が直感的で業務効率が大幅にアップしました。		
2023/2/4	C035	医療機関	100名未満	一般職	男性	30代	初期費用が高く、コスト面で負担を感じます。		
2023/2/5	C036	IT業	100名未満	役員・経営	男性	30代	導入が容易でスピーディに使用開始が可能でした。		

+ テンプレート作成

コードインタープリター

アプリディレクター2502

テキストマイニングによるVOC
:: (顧客の声) 分析 ※ファイル添付必須

データ前処理・クレンジング75% ※添付ファイル必須

添付ファイル→PPT化

テキストマイニングによるVOC (顧客の声) 分析 ※ファイル添付必須

分析対象のファイル形式

例：csv、xlsx、csv UTF-8(コンマ区切り) など

(任意)分析目的

例：製品満足度の評価、サービス改善のための課題抽出、ブランドイメージの評価、新製品開発のためのニーズ調査 等

1. レポート概要 (Executive Summary)

- 分析の背景・目的
分析を行う理由や狙い、{顧客から見たexaBase 生成AIの課題の把握}など。
- 分析対象(期間・データソースなど)
VOCデータの概要(収集期間、データ件数、情報源 など)。
データの前処理(クリーニング手順、テキストの正規化 など)を簡潔に記載。
- 主な成果と推奨アクション

アンケート、口コミを多面的に分析

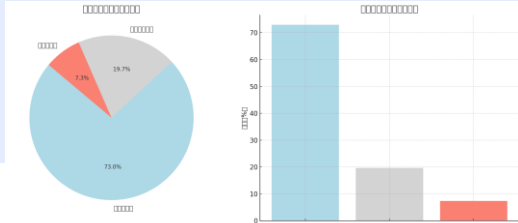
複数の手法による分析結果の出力

辞書ベースの感情分析を行った結果、VOCデータ全体の感情分布は以下のようになりました：

- **ポジティブ**: 73.0%
- **ニュートラル**: 19.67%
- **ネガティブ**: 7.33%

ポジティブな意見が大多数を占めていることが分かります。一方で、ネガティブな意見が約7%含まれており、この顧客層の課題を特定して改善することが重要です。

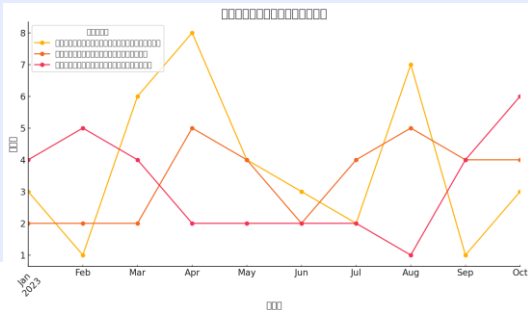
次に、この感情分析結果を円グラフや棒グラフで視覚化し、



上記の時系列グラフは、上位3つのキーワードの月別出現数の推移を示しています。

- 各キーワードの出現状況や変動パターンを確認できます。
- 特に特定の月に急激な増減が見られる場合、それに関連したイベント（キャンペーン、トラブルなど）が影響しているかもしれません。

次に、「共起分析」と「共起ネットワーク可視化」を行い、キーワード同士の関連性を明らかにします。この部分を進めてよろしいでしょうか？



レポートの出力

1. レポート概要 (Executive Summary)

背景・目的

本分析では顧客からのVOC（Voice of Customer）データを解析し、exaBase生成AIに関する顧客の不満点や満足点を正確に把握することを目的としています。本レポートは、感情分析、キーワード頻度、共起関係、トピックモデリングなど、多角的な手法を用いてVOCを深掘りしました。

分析対象

- **収集期間**: 2023年1月1日～2023年10月27日
- **データ件数**: 300件
- **情報源**: 顧客コメントテキスト
- **主要項目**: 性別、年代、業種、会社規模、役職、自由形式コメント

主な成果と推奨アクション

主な成果

- 感情分布**
顧客の意見の73%がポジティブで、7%程度のネガティブ意見が目立ちます。主にコストと安定性に関する課題がネガティブコメントに関連していました。
- 頻出キーワードと出現パターン**
「柔軟なカスタマイズ」「導入の容易さ」「コストパフォーマンス」が頻繁に言及される一方、「初期費用」や「サポートの遅延」など課題点も確認されました。
- トピックモデリング**
VOCを5つの主要なテーマに分類し、特に「コスト/導入プロセス」と「安定性・サポート」に関する声が強調されました。
- 時系列推移**
ある時期に特定のキーワード（例: 新機能追加）が急増していることが確認されました。これらは特定のイベントや製品アップデートに起因するものと考えられます。

過去データから予測モデルを作り、実行

過去3年分の売上データを準備

Date	Store_ID	Sales	Customers	Weather	Campaign	Holiday_Flag
2022/1/1	Store_001	3813341	447	晴れ	チラシ配布	1
2022/1/2	Store_001	3962199	411	雪	ポイント2倍	1
2022/1/3	Store_001	2251286	224	雪	クーポン配布	0
2022/1/4	Store_001	3548128	356	雪	チラシ配布	0
2022/1/5	Store_001	4246093	408	雨	ポイント2倍	0
2022/1/6	Store_001	4534889	301	晴れ	チラシ配布	0
2022/1/7	Store_001	2103440	453	雨	クーポン配布	0
2022/1/8	Store_001	2181661	316	晴れ	チラシ配布	1
2022/1/9	Store_001	2459316	227	曇り	チラシ配布	1
2022/1/10	Store_001	2143824	354	雨	クーポン配布	0
2022/1/11	Store_001	2542666	268	晴れ	チラシ配布	0
2022/1/12	Store_001	2434020	424	晴れ	ポイント2倍	0
2022/1/13	Store_001	3120254	272	曇り	チラシ配布	0
2022/1/14	Store_001	4472794	274	雪	チラシ配布	0
2022/1/15	Store_001	3302643	228	雪	なし	1
2022/1/16	Store_001	3759604	410	曇り	ポイント2倍	1
2022/1/17	Store_001	2543183	398	雪	なし	0
2022/1/18	Store_001	2445743	271	晴れ	チラシ配布	0
2022/1/19	Store_001	3712761	281	雨	チラシ配布	0
2022/1/20	Store_001	4610282	469	雪	チラシ配布	0
2022/1/21	Store_001	4094451	379	曇り	なし	0
2022/1/22	Store_001	4750171	473	晴れ	クーポン配布	1
2022/1/23	Store_001	2331058	275	曇り	なし	1

データを添付し売上予測モデル作成を依頼

添付のファイル名:storesales_1000sample.xlsxはある店舗の1000店舗の2022年1月1日から2024年12月31日までの売上実績のデータです。
このデータを読み込み、以降の2025年以降の売上予測modelをつくってください。
念のため、このファイルデータの2024年12月データはテスト用としてモデル作成には使わず、テストを行い、出力形式に従いモデルの精度を表示してください。
精度は実績からズレを%で表示してください。

出力形式

日付	売上予測	実績	精度
2022-12-01			

storesales_1000sample.xlsx
ファイル

↓

精度の高いモデル
ができた

2024年12月の予測結果と実績、および精度は以下の通りです：

日付	売上予測	実績	精度 (%)
2024-12	105,869,900	98,604,195	7.37

SARIMAモデルで、2024年12月の売上予測と実績との誤差率（ズレ）は約 **7.37%** です。次に、このモデルを用いて2025年以降の売上を予測します。

過去データから予測モデルを作り、実行

モデルを活用し売上予測を実行

⚡ 再調整したSARIMAモデルによる2025年の月次売上予測結果は以下のようになりました（予測は月末日時点の金額です）：

日付	売上予測
2025-01-31	110,311,100
2025-02-28	109,242,600
2025-03-31	95,516,910
2025-04-30	115,565,200
2025-05-31	109,163,200
2025-06-30	113,037,400
2025-07-31	105,137,800
2025-08-31	105,233,600
2025-09-30	107,569,100
2025-10-31	109,645,000
2025-11-30	111,466,600
2025-12-31	105,626,100

モデル調整により、2月以降の急激な伸びが抑えられ、より安定した予測が得られました。

この結果を確認し、さらに外部要因を含めたモデル改善を行う必要があればお知らせください。また、グラフ描画を行い、新しい予測と実績の比較を視覚化することも可能です。どのように進めるかご指示ください。

別のモデルを構築

⚡ グラフでは、再調整されたSARIMAモデルによる予測（点線）と実際の売上データ（実線）が示されています。2月以降の不自然な値が修正され、全体的に実績に近い予測結果となっています。

改善点

- モデル調整後、季節性とトレンドがより現実的なものとなり、売上予測の急激な伸びが抑えられました。
- 予測値の変動は実績の範囲内に収まり、より信頼性の高い結果が得られています。

次に進む必要がある場合、お知らせください。たとえば、さらなる改善、別のモデル（LSTMなど）の試行、または追加の外部変数（キャンペーン効果や天気など）の考慮なども可能です。

提案

現状の環境では、時系列モデル（ARIMAやSARIMA）をさらに調整し、性能向上を試みるか、別の機械学習を用いたアプローチ（XGBoostやRandom Forestなど）で売上予測を行うことができます。これらを試してみませんか？



👤 XGBoostでお願いいたします

⚡ XGBoost（Extreme Gradient Boosting）は、回帰・分類問題に広く使用される強力な機械学習手法です。XGBoostを用いて売上予測モデルを構築します。以下の手順を実行します。

手順

1. データ前処理

- 特徴量を構築（例：日付要素など）。

20人の従業員を休みや職能、勤務制限を配慮して配置するシフト作成

LLMと何度も対話し作成したプロンプト

```
あなたはPythonを用いたシフトスケジュール最適化のエンジニアです。添付のExcelデータを基に、
全体バランスと各ルールを遵守しながら効率的にシフトスケジュールを作成してください。

---

## 【入力データ】（添付のExcelファイルから取得）

### ■ Excelファイルのシート構成
添付されたExcelファイルには以下のシートが含まれています。それぞれのシートからデータを取得してください。

1. **従業員リスト** (Employees シート)
  - A列: 従業員名

2. **日数リスト (各コマ)** (Days シート)
  - A列: コマ名 (例: 「1日目1コマ」)

3. **シフト種別** (Shift_Types シート)
  - A列: シフト名 (例: 「ホール」 「カウンター」)

4. **各シフトごとの日ごとの需要** (Shift_Demand シート)
  - A列: コマ名 (Days シートと一致)
  - B列: ホールの必要人数
  - C列: カウンターの必要人数

5. **従業員ごとの休日希望** (Holiday_Preferences シート)
  - A列: 従業員名
  - B列: コマ名 (Days シートと一致)

6. **従業員ごとのシフト希望回数** (Shift_Preferences シート)
  - A列: 従業員名
  - B列: 希望回数

7. **連続シフト数上限** (Constraints シート)
  - A1セル: 連続シフト数の上限値

8. **従業員ごとの対応可能シフト** (Employee_Shifts シート)
  - A列: 従業員名
  - B列以降: 対応可能シフト (「ホール」 「カウンター」のいずれか)

---

## 【制約条件】

1. **同一コマの重複禁止**
  - **同一日・同一コマで1人が複数のシフトに割り当てられないことを厳密にチェックする。**
  - 最終出力前に、この条件を満たしているかを確認し、**違反があった場合は再割り当てを行う**

2. **全シフト枠の充足**
  - 各コマの需要 (ホール、カウンターの必要人数) は Shift_Demand シートの情報を満たす。
```

指示通りのシフトが組めない ×

⚡ シフトスケジュール表:

同じコマに2つのシフトが割り振られている

時間	ホール	カウンター
1日目1コマ	今中社員, 佐藤, 加藤, 大畑, 山田主任	今中社員
1日目2コマ	今中社員, 佐藤, 加藤, 大畑, 山田主任, 川田	今中社員 佐藤
1日目3コマ	今中社員, 佐藤, 加藤, 大畑, 山田主任, 川田	今中社員 佐藤
2日目1コマ	今中社員, 佐藤, 加藤, 大畑, 山田主任	今中社員
2日目2コマ	今中社員, 佐藤, 加藤, 大畑, 山田主任, 川田	加藤, 大畑
2日目3コマ	加藤, 大畑, 山田主任, 川田, 後藤, 本田	安達, 山田主任
3日目1コマ	山田主任, 川田, 後藤, 本田, 田中店長	大谷
3日目2コマ	山田主任, 川田, 後藤, 本田, 田中店長, 石山	大谷, 安達
3日目3コマ	川田, 後藤, 本田, 田中店長, 石山, 石田	安達, 川田
4日目1コマ	後藤, 本田, 田中店長, 石山, 石田	大谷
4日目2コマ	後藤, 本田, 田中店長, 石山, 石田, 菅野社員	大谷, 安達

1問ずつLLMから質問してもらう形式で進めると上手くいくケースもある

一問一答形式でシフト作成をすすめる

パチンコホールの勤務シフトを作成したいです。必要な項目を私に1つずつ質問する形で作成してください。

分かりました！パチンコホールの勤務シフトを作成するには、いくつかの基本的な情報が必要です。以下質問形式で進めていきますのでご回答ください。

質問1：

まず、何日間分のシフトを作成しますか？
(例：1週間、1ヶ月など)



7日分です。また1日あたり3コマあり、合計21コマ分のシフトとなります

ありがとうございます！1週間（7日分）で、1日あたり3コマ、合計21コマのシフトを作成しますね。

質問2：

1日の3コマそれぞれの時間帯を教えてください。
(例：朝：9:00～12:00、昼：12:00～18:00、夜：18:00～23:00)



リクエストに近いシフトができる

以下は、シフトの21コマ分を表形式でまとめたものです：

日	コマ	ホール担当メンバー	カウンター担当メンバー
1日目	1コマ目	大畑, 後藤, 今中社員, 松田, 菅野社員	川田
1日目	2コマ目	松田, 安達, 長谷川, 藤田, 山田主任, 佐藤	菅野社員, 松宮
1日目	3コマ目	大畑, 後藤, 菅野社員, 加藤, 本田, 田中店長	松田, 安達
2日目	1コマ目	後藤, 菅野社員, 田中店長, 大谷, 馬場	大畑
2日目	2コマ目	松田, 川田, 安達, 長谷川, 加藤, 馬場	山田主任, 石田
2日目	3コマ目	安達, 長谷川, 佐藤, 加藤, 本田, 田中店長	大畑, 松田
3日目	1コマ目	川田, 安達, 佐藤, 田中店長, 馬場	菅野社員
3日目	2コマ目	大畑, 後藤, 川田, 藤田, 田中店長, 大谷	松田, 本田
3日目	3コマ目	川田, 長谷川, 加藤, 田中店長, 馬場, 石田	山田主任, 佐藤
4日目	1コマ目	後藤, 今中社員, 菅野社員, 馬場, 石田	山田主任
4日目	2コマ目	今中社員, 松田, 藤田, 山田主任, 田中店長, 鈴木社員	馬場, 石田
4日目	3コマ目	今中社員, 川田, 安達, 佐藤, 鈴木社員, 石山	後藤, 松田



EXAWIZARDS

AIで、ひとに力を。よりよい社会を。