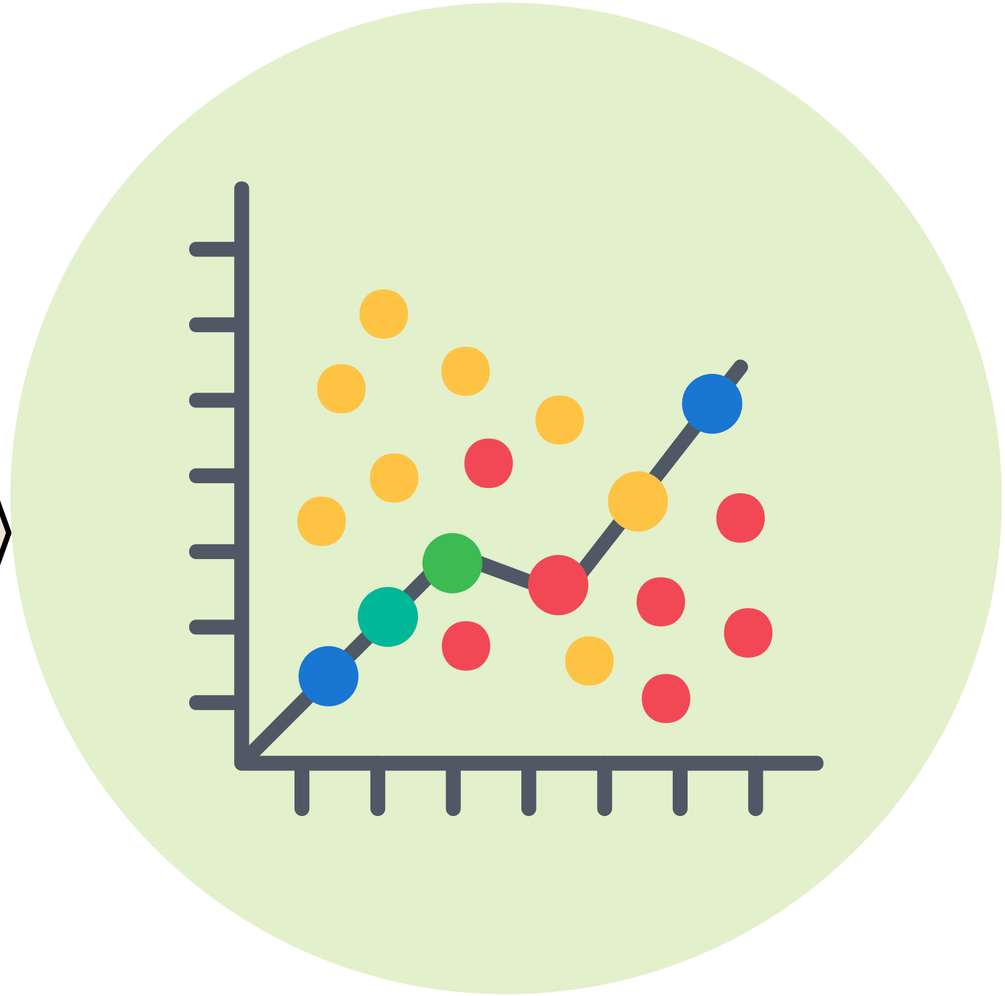
python™

Statsmodels

Linear Mixed Model

-by:
-Rishabh Singh
-Data Scientist



Learning Outcomes

Modeling Hierarchical and Longitudinal Data with Linear Mixed Models

Leveraging statsmodels' MixedLM in Python

Plot graphs of linear mixed model

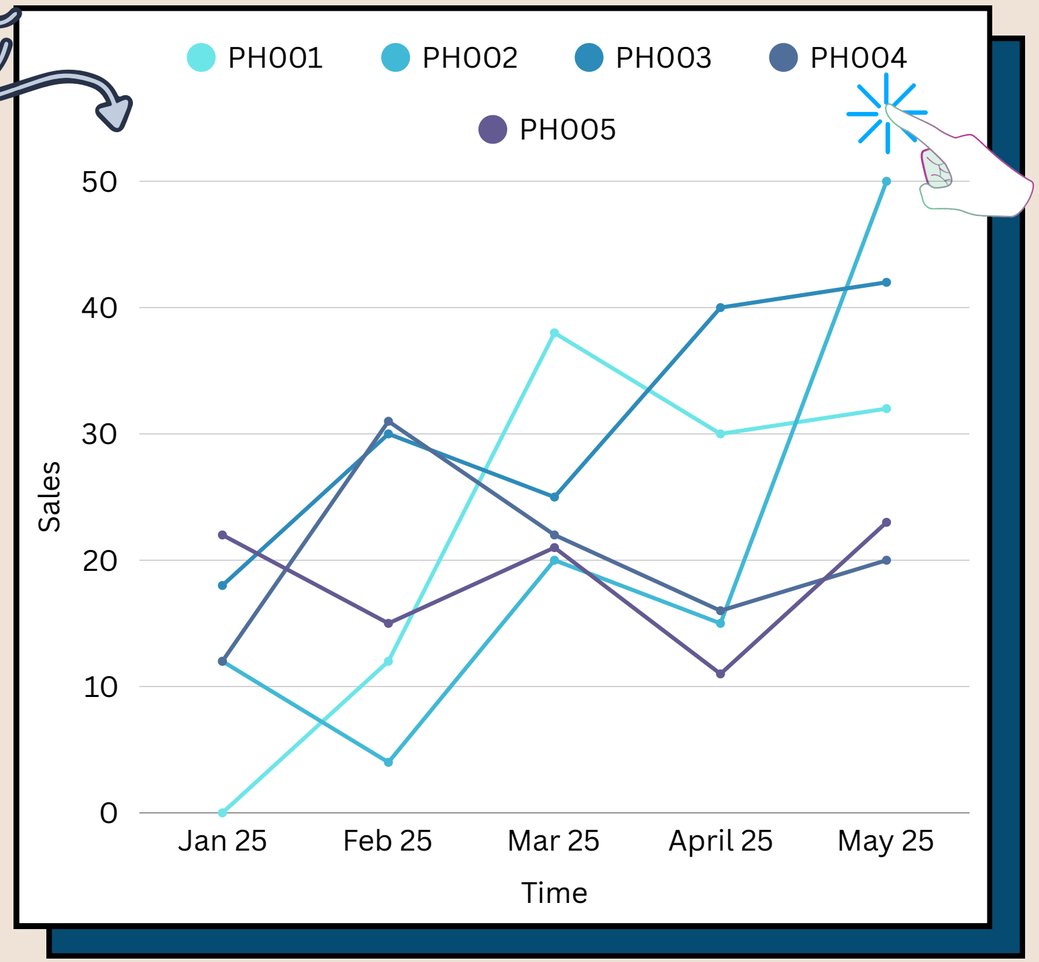
Agenda

- Introduction to Hierarchical and Longitudinal Data
- Overview of Linear Mixed Models (LMMs)
- Implementing LMMs with statsmodels' MixedLM
- Model Diagnostics and Interpretation
- Case Study: Pharmacy’s Promotion Impact on Sales
- Resources and Further Reading

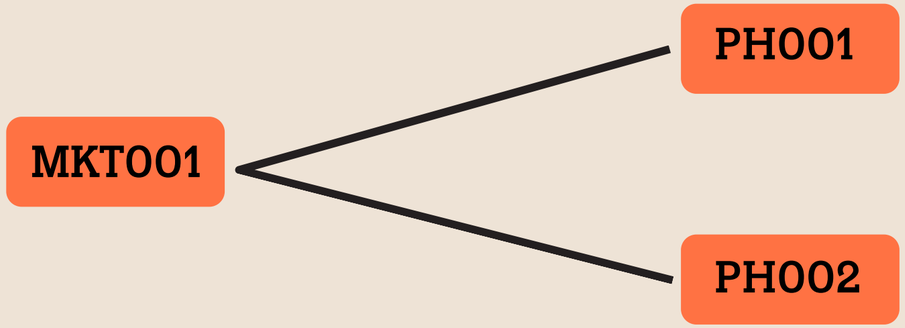
Longitudinal Data

click here

Longitudinal data involves repeated observations of the same subjects over time, allowing for the analysis of temporal changes within those subjects.



Hierarchical Data

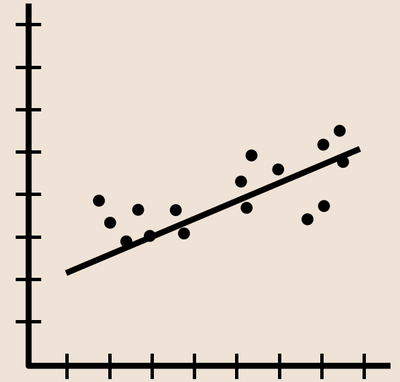


Hierarchical data is structured in a tree-like format, where each record (node) has a parent-child relationship. This model organizes data into multiple levels, resembling an organizational chart or directory structure.

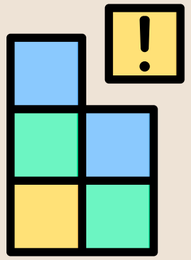
Market Code	Pharmacy Code	Month	Sales Amount (₹)
MKT001	PH001	Jan	1,200
MKT001	PH001	Feb	1,150
MKT001	PH002	Jan	1,100
MKT001	PH002	Feb	1,050
MKT002	PH003	Jan	1,300
MKT002	PH003	Feb	1,250

click Here

Limitations of Traditional Linear Models

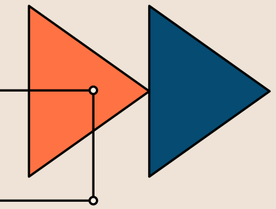


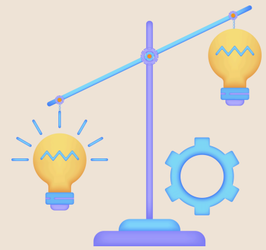
1. Assumption of Linearity
   - **Presumes** a straight-line relationship between variables.
   - Fails to capture **complex, nonlinear patterns** in real-world data.
2. Sensitivity to Outliers
   - **Outliers** can disproportionately **influence the regression line**.
   - Leads to **skewed results** and reduced predictive accuracy.



3. Homoscedasticity Requirement
   - **Assumes constant variance** of errors across all levels of independent variables.

4. Assumption of **Independence**
   - **Assumes** observations are **independent of each other**.
   - Often violated in time-series or hierarchical data, leading to **biased estimates**.





Introduction to Linear Mixed Models (LMMs)

Linear Mixed Models are an extension of linear regression that incorporate both fixed and random effects. They are particularly useful for analyzing data with hierarchical or grouped structures, such as repeated measures or longitudinal studies.

Key Components:

- **Fixed Effects:** Parameters associated with the entire Data or Market Level (e.g., overall effect of drugs).
- **Random Effects:** Random variables associated with individual units or Pharmacy from a population (e.g., variability among different pharmacies).

Mathematical Formulation of Linear Mixed Models (LMMs)

$$y = X\beta + Zb + \varepsilon$$

- **Observed Response (y)**: The dependent variable being modeled.
- **Fixed Effects (Xβ)**: Represent the population-level effects, such as overall trends or treatment effects.
- **Random Effects (Zb)**: Capture individual-specific deviations, accounting for variability across groups or subjects.
- **Residual Errors (ε)**: Represent unexplained variability, assumed to be normally distributed.

Implementing Linear Mixed Models (LMMs) with statsmodels

```python
import statsmodels.formula.api as smf

# Define and fit the Linear Mixed Model
model = smf.mixedlm("Sales ~ Adstock_Promotion + Lag1_Sales + Lag2_Sales",
        df_model, groups=df_model["PharmacyID"],
re_formula="~Adstock_Promotion")
result = model.fit()

# Output the model summary
print(result.summary())
```

Sales, Adstock, Lag1, Lag2
are
Feature/Columns from
dataframe

Implementing Linear Mixed Models (LMMs) with statsmodels

- **Intercept** (50.123): Baseline sales
- **Adstock_Promotion** (2.345): A one-unit increase in adstock-transformed
- **Lag1_Sales** (0.678): Previous sales positively influence
- **Group Var** (4.567): Variance in sales across pharmacies

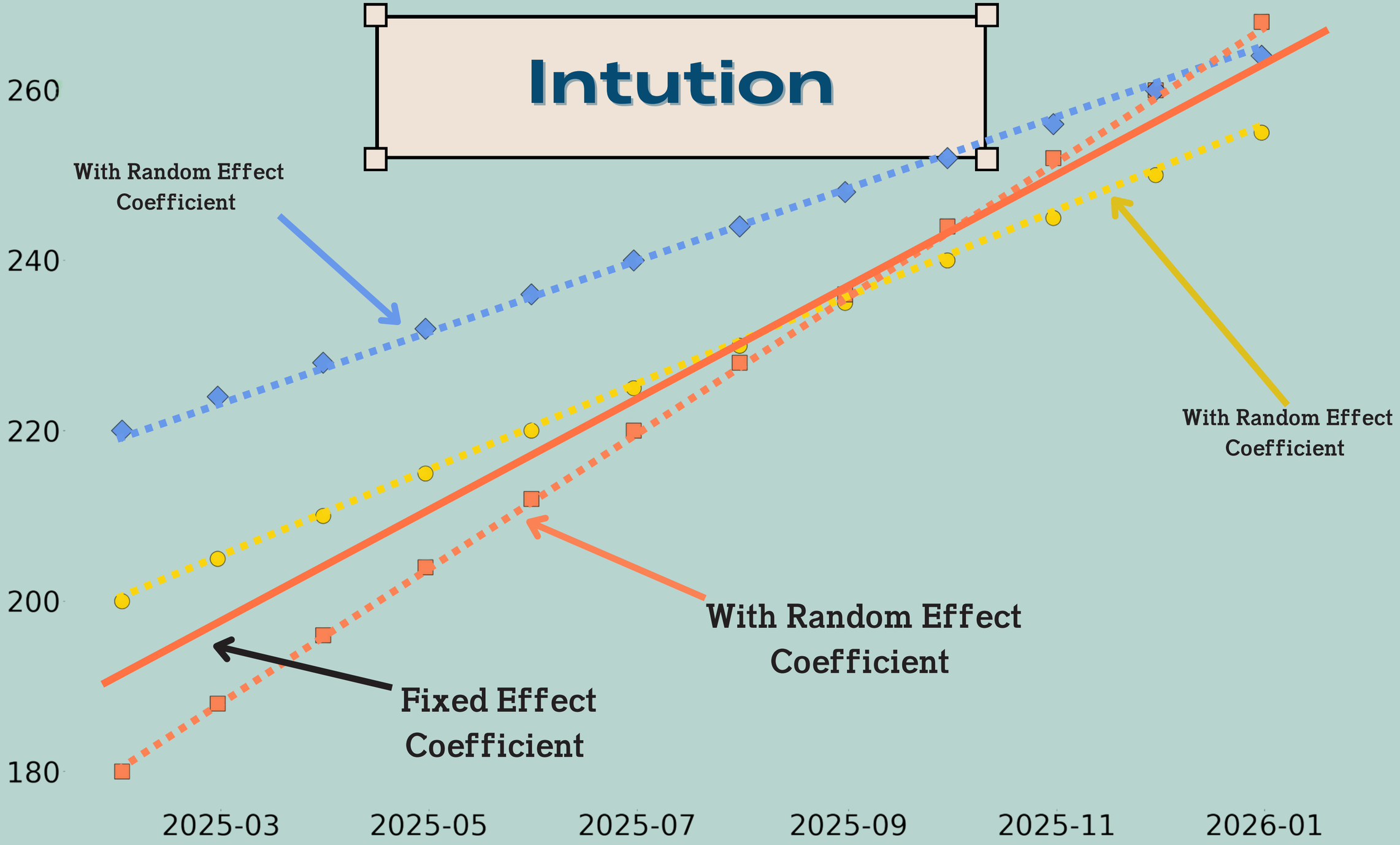
```
Mixed Linear Model Regression Results
============================================================
Model:             MixedLM Dependent Variable: Sales
No. Observations:  1,200   Method:             REML
No. Groups:        100     Scale:              8.4321
Min. group size:   12      Log-Likelihood:     -3,456.789
Max. group size:   12      Converged:          Yes
Mean group size:   12.0
------------------------------------------------------------
                  Coef.  Std.Err.    z    P>|z| [0.025 0.975]
------------------------------------------------------------
Intercept        50.123    1.234  40.620 0.000 47.704 52.542
Adstock_Promotion 2.345    0.456   5.144 0.000  1.451  3.239
Lag1_Sales        0.678    0.089   7.618 0.000  0.503  0.853
Lag2_Sales        0.321    0.076   4.224 0.000  0.172  0.470
Group Var         4.567    0.789
============================================================
```

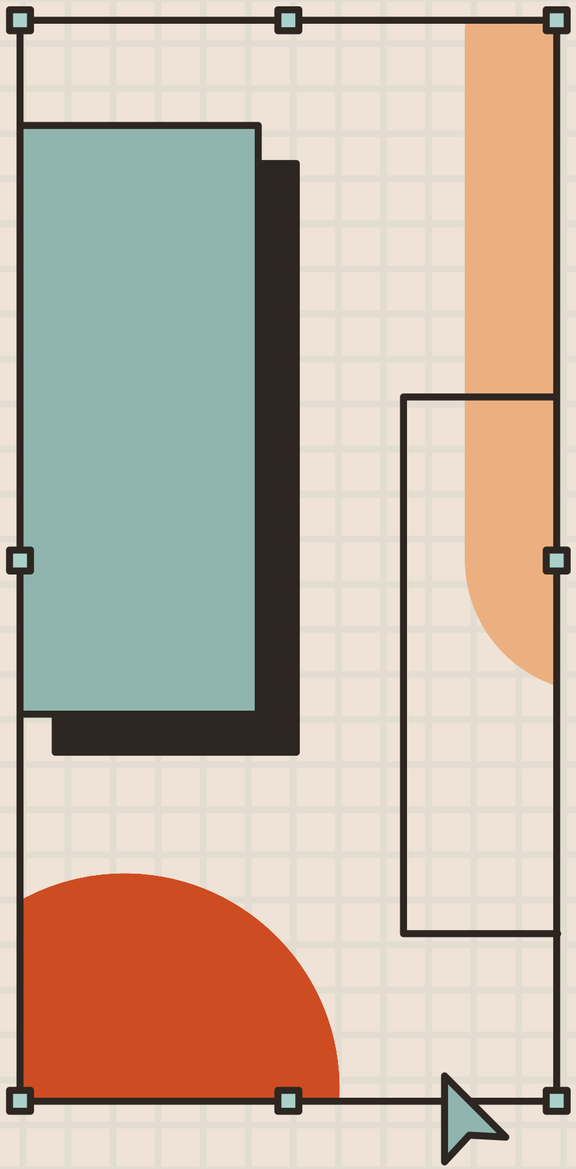
- **Observations**: 1,200
- **Groups**: 100 pharmacies
- **Scale** (8.4321): Residual variance, representing unexplained variability.
- **Log-Likelihood** (-3,456.789): Measure of model fit; higher values indicate better fit.
- **Converged**: Yes

Implementing Linear Mixed Models (LMMs) with statsmodels

```
random_effects = result.random_effects


_____________________________________________________________

Output :-

  {
    'Pharmacy_1': {'Intercept': 2.5, 'Adstock_Promotion': 0.3},
    'Pharmacy_2': {'Intercept': -1.2, 'Adstock_Promotion':
0.5},..
    }
```

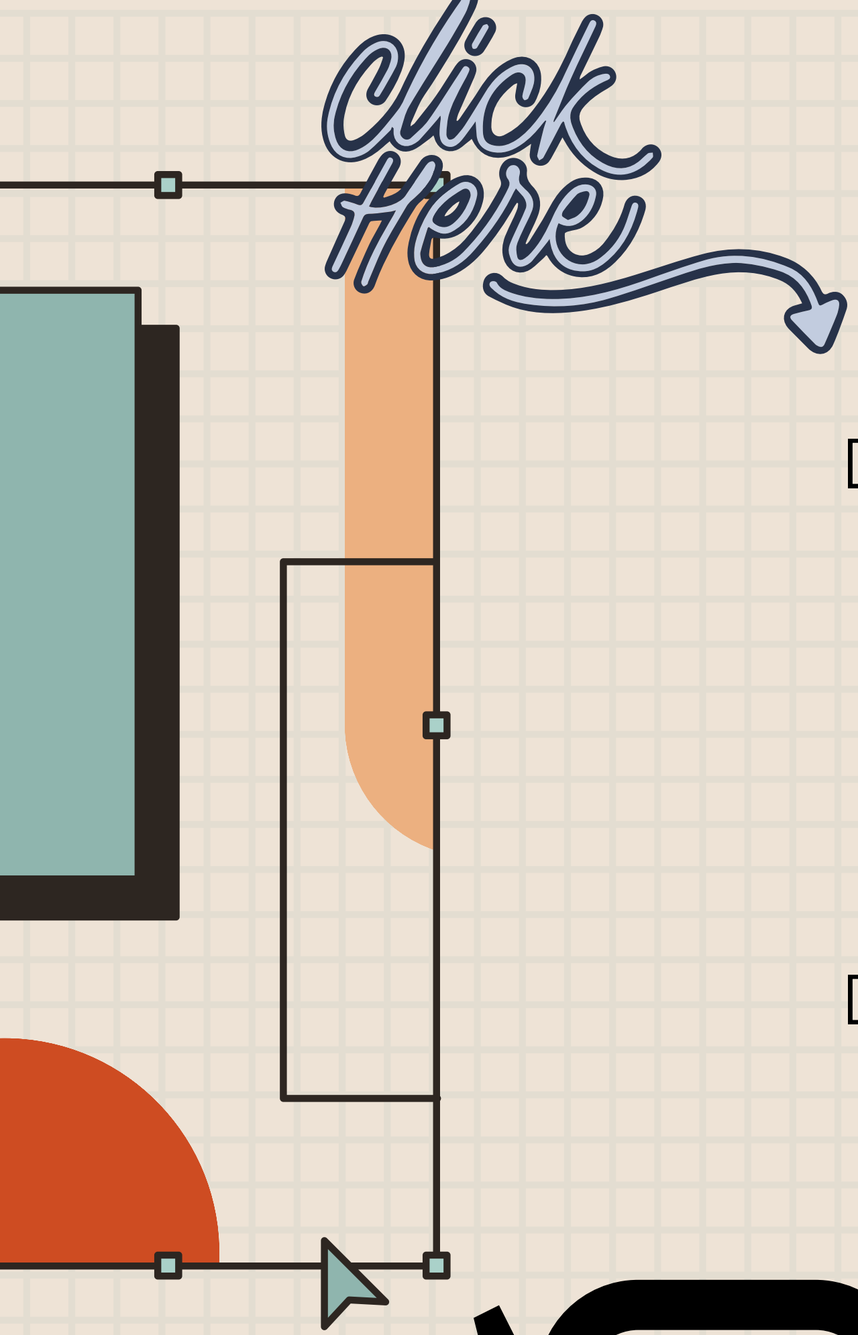
Intution


260
240
220
200
180
2025-03
2025-05
2025-07
2025-09
2025-11
2026-01
With Random Effect Coefficient
With Random Effect Coefficient
With Random Effect Coefficient
Fixed Effect Coefficient



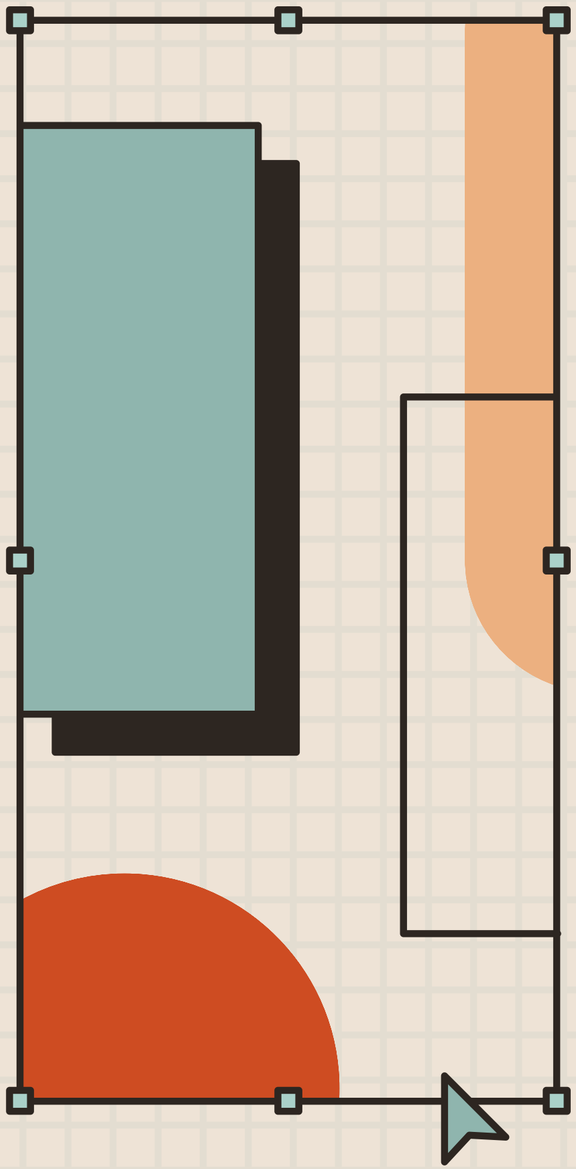
click Here

Example for MixedLM



Resources and Further Reading

- https://stats.oarc.ucla.edu/introduction-to-linear-mixed-models/
- Statsmodels.regression.mixed_linear_model.MixedLM
- https://github/statsmodels/regression/mixed_linear_model



Any Questions?

The End