Ryder®
Ever better.™

Customer No-Show Prediction

Liangzhu Chen, Hsiao Wei Liu, Michael Liu, Yanmin Luo
Yanke Sun, Xinyu Wang, Jason Xu, Jiaxin Xu, Aixin Zhou

Agenda

- Team
- Executive Summary
- Project Background
- Data Summary
- Analytics
- Insights and Recommendations
- Q&A

Team

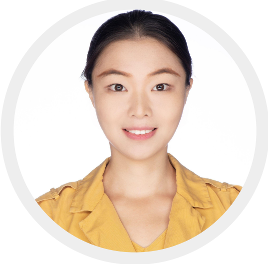
Hsiao-Wei Liu
hl3663@columbia.edu
Data Analyst
May 2024

Jiaxin Xu
jx2530@columbia.edu
Business Analyst
May 2024

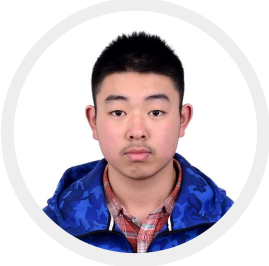
Jason Xu
rx2223@columbia.edu
Business Analyst
May 2024

Michael Liu
sl5300@columbia.edu
Financial Analyst
May 2024

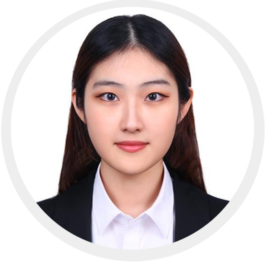
Liangzhu Chen
lc3694@columbia.edu
Financial analyst
Dec 2023



Yanmin Luo
yl5292@columbia.edu
Marketing Analyst
May 2024



Yanke Sun
ys3686@columbia.edu
Data analyst
May 2024



Xinyu Wang
xw2904@columbia.edu
Data Analyst
Dec 2023

Aixin Zhou
az2767@columbia.edu
Data Analyst
May 2024

Executive Summary

Project Goal

Develop a **customer maintenance appointment no-show rate predictive solution** that helps minimize customer no-show rate

Deliverables

1. **Model** to be integrated on Ops Platform
2. **Data visualizations** that yield business insights

Analytical Approach

Random forest model trained on 80% of 100,000 appointment records that underwent stratified sampling

Insights and recommendations

1. Review resource allocation
2. Customer engagement

Project Background

Challenge

- Ryder is trying to **minimize customer no-shows** in maintenance appointments.

Proposed Solutions

- We aim to develop a **predictive model** that forecasts customer show-up rates for 21 time slots over a week, covering 3 daily shifts.

Expected Benefits

1. **Improve** Ryder's **current operation efficiency** by suggesting scheduling time slots with the highest likelihood of success.
2. **Increase customer satisfaction** by **reducing the number of calls** with customers to schedule maintenance events.
3. The **predictive model** can be further developed to **integrate with Ryder's auto scheduling platform** to suggest **optimal** maintenance time slots for individual vehicle/customer.
4. **Reduce administrative cost** associated with current scheduling process once model is integrated into auto scheduling.

Data Summary

Data Dimensions
1.4M observations
115 features

Attributes
- Customer
- Appointment
- Fleet
- Geographic

Data Types
- Date
- Categorical
- Numeric
- Binary

Distribution
- ~20% appointments
- ~80% drive ups

Analytics

- Feature Engineering
- Random Forest Model
- Feature Importance Analysis

Feature Engineering

Goal: Convert raw data into model inputs

1. **Gather Data**
   Combine initial datasets from 2018 to 2020

2. **Features Selection**

```
columns_used = [
    'VEHICLE_NBR',
    'CUST_ACCOUNT_NBR',
    'APPOINTMENT_DT',
    'UNIT_CHECKIN_DT',
    'APPOINTMENT_PRIORITY_ID',
    'APPOINTMENT_TYPE_ID',
    'ISSUING_REG_DESC',
    'INDUSTRY_GROUP',
    'PM_TYPE',
    'RIDE_VEHICLE_TYPE_DESC'
]
```

not used as final model inputs

3. **Handle Missing Values (Dropped)**
   Datetimes, Priority ID, PM type

4. **Convert Data Types (E.g. integer to string)**

5. **Create Additional Datetime Features**
   Weekday
   Time period during the day

Feature Engineering (Cont.)

6. One hot encoding

	PM_TYPE		PM_TYPE_BPM	PM_TYPE_APU	PM_TYPE_LOAD
Row 1	BPM		1	0	0
Row 2	APU	→	0	1	0
Row 3	LOAD		0	0	1

7. Labeling ground truth

If appointment and check-in:

- Dates are the same
- Weekdays are the same
- Time period are the same

→ Assign 1

Else (not the same) → Assign 0

Feature Engineering (Cont.)

8. Stratified sampling

All Data (~1,400,000 obs)

All Appointment Data
(~250,000 obs)

Show Up
(50,000 obs)

No Show
(50,000 obs)

Final input data

X (Independent variables)
- Customer info (dummy variables)
- Appointment weekday and time period (dummy variables)

Y (Dependent variable)
- Column "SHOW_UP"
- Stored as 1 and 0 (1,0,0, 1, …)

Model - Random Forest

Hyperparameter Tuning

- ❑ *n_estimators* → number of decision trees
- ❑ *max_depth* → maximum levels within each tree
- ❑ *min_samples_split* → minimum data size for each split between levels

Random walk to get the best **AUC**

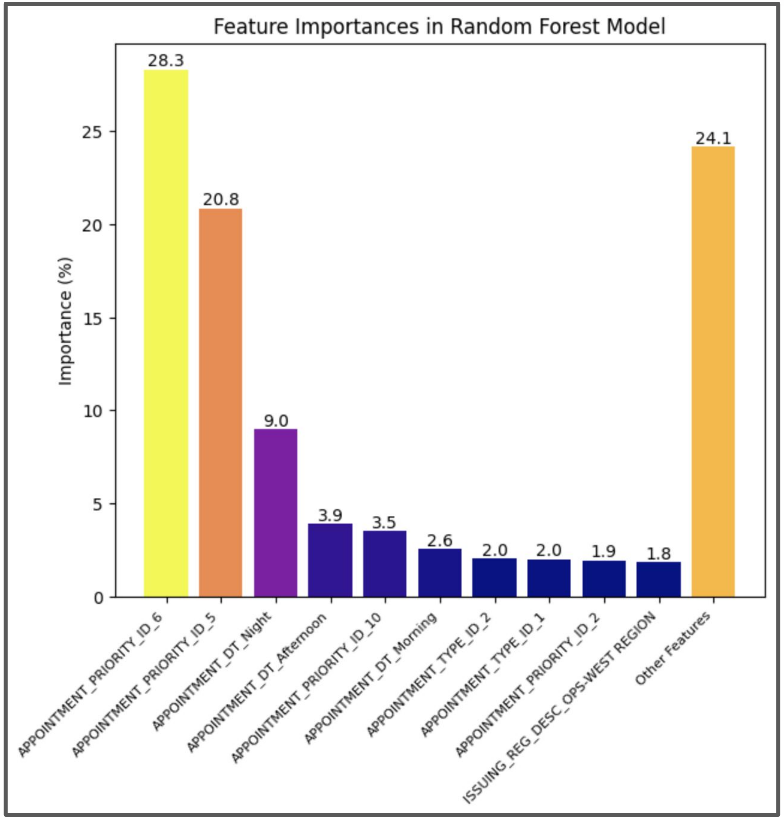
Train and Test

- ❑ Train → 80,000 obs
- ❑ Test → 20,000 obs
- ❑ Drop ID and all attributes associated with check-in prior to fitting

Evaluation of Model

1. Train
   - **AUC: 88.28%** Accuracy: 80%
   - Non show-up
     - Precision: 85%
     - Recall: 73%
   - Show-up
     - Precision: 77%
     - Recall: 87%
2. Test
   - **AUC: 85.02%** Accuracy: 78%
   - Non show-up
     - Precision: 82%
     - Recall: 75%
   - Show-up
     - Precision: 71%
     - Recall: 85%

Feature Importance Analysis



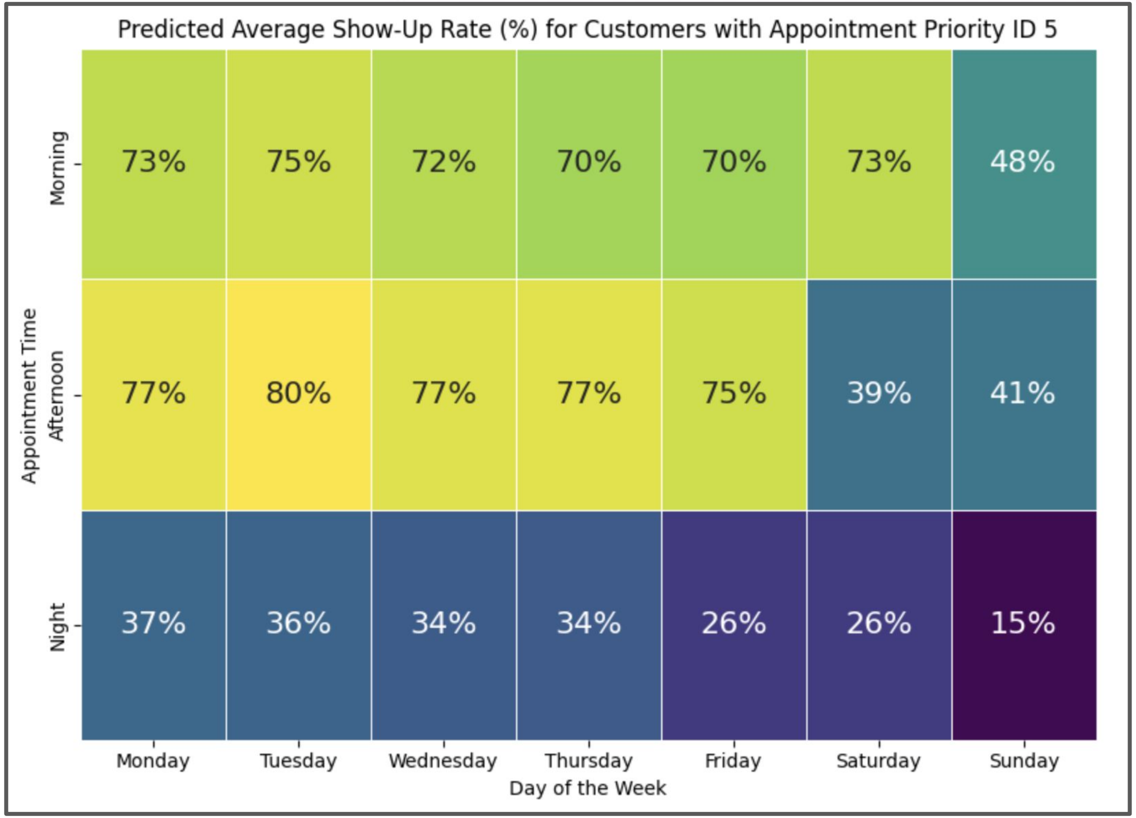
Appointment Priority IDs 5 (hard schedule) and 6 (maintenance with no vehicle substitution) are the most important features, taking up about half of the model's total feature importance.

Nighttime appointments have a stronger influence in shaping the model's outcome than morning / afternoon.

Apart from the top 10 features, all other features (64 in total) represent about **¼ of the total feature importance**.

Business Use Case

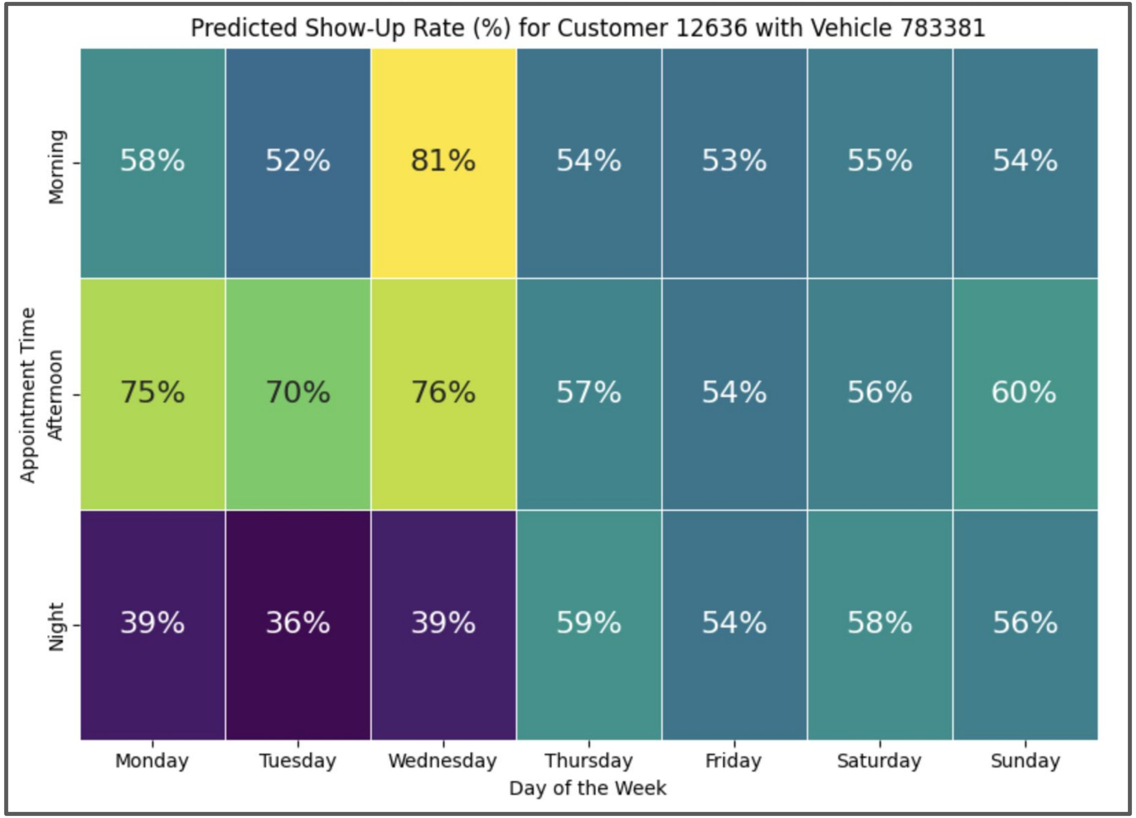
Predict by Segments



To gain insights for a specific customer segment, we first use the model to predict the show-up probability for all appointments. Then we filter these predictions based on selected criteria and calculate the average show-up rate for each time slot.

Alternatively, we could predict binary outcomes (show-ups and no-shows) for each case and compute the proportion of show-ups in each time slot.

Predict by Customer / Vehicle ID Combination

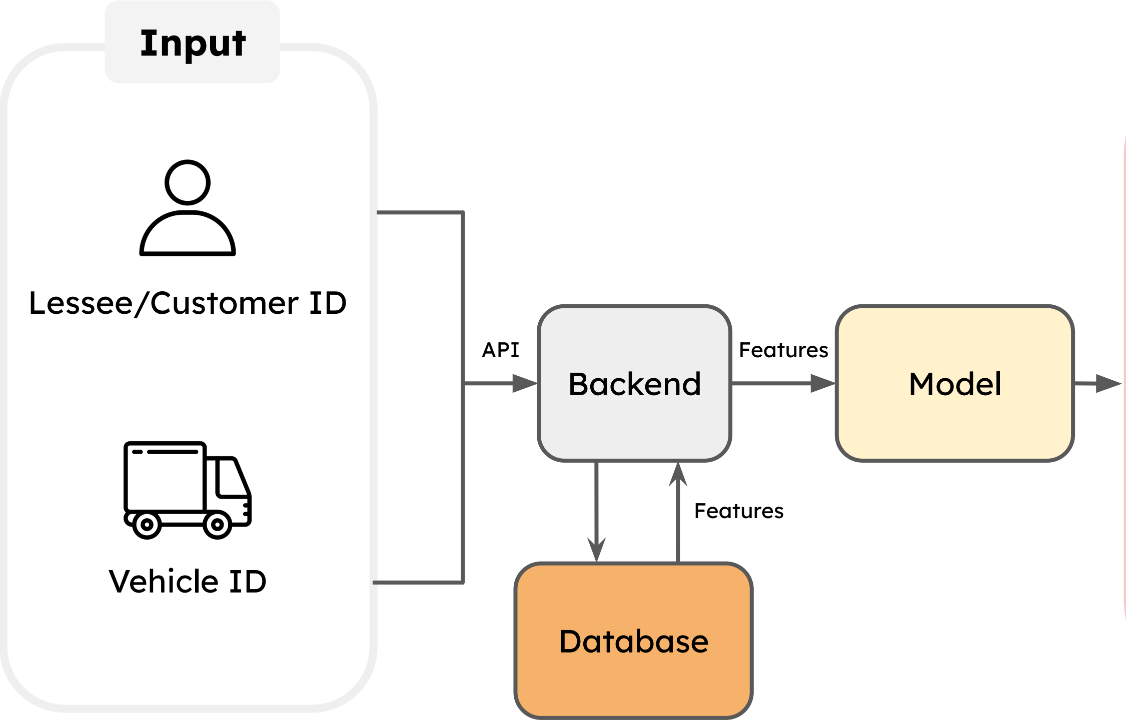


To predict individual customer/vehicle show-up rates, we'll first select specific customer/vehicle ID combinations from all appointment data.

Next, we'll remove the appointment date time features and create 21 time slots, duplicating each customer's remaining features for these slots.

Finally, we'll use the same prediction mechanism to estimate the show-up rate for each customer across these time slots.

Implementation Recommendation



Output

Predicted show-up probabilities of each customer/vehicle combination in 21 time slots

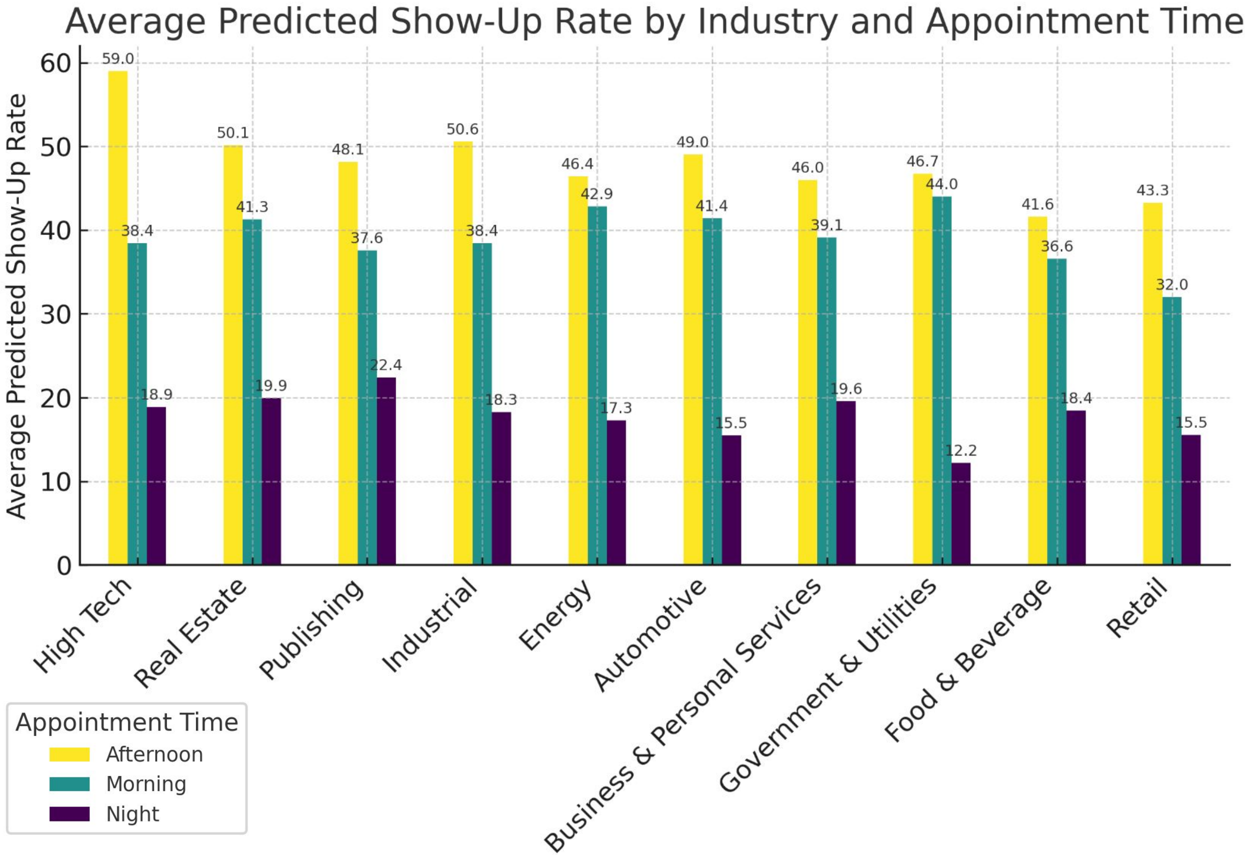
Days	Morning	Afternoon	Night
Monday	P1	P2	P3
Tuesday	P4	P5	P6
Wednesday	P7	P8	P9
Thursday	P10	P11	P12
Friday	P13	P14	P15
Saturday	P16	P17	P18
Sunday	P19	P20	P21

The scheduling platform can automatically assign user appointments to time slots with predicted show-up rates above a set threshold.

Business Insights & Recommendations

- Industries
- Appointment Priority IDs
- Vehicle Types
- Regions

Industries



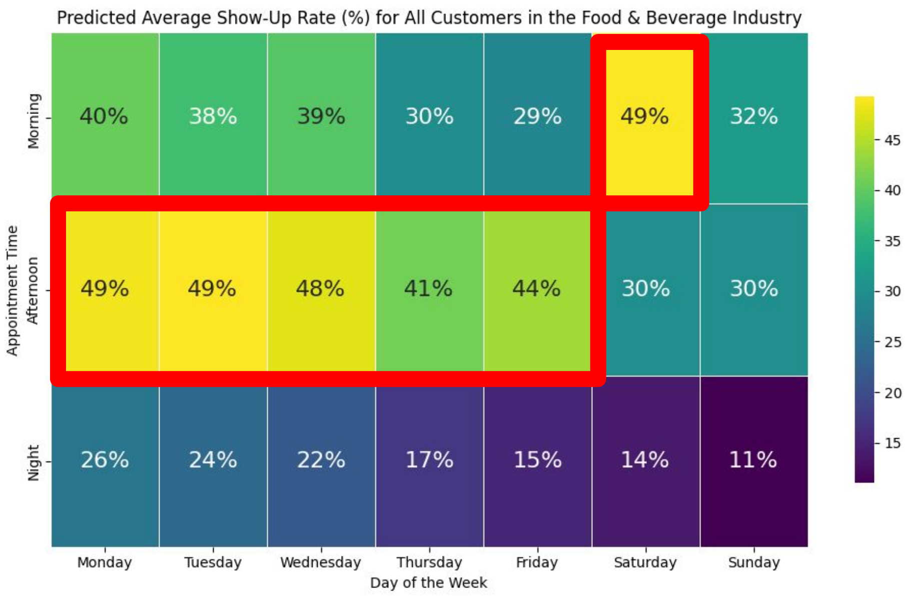
There's no significant difference across industries.

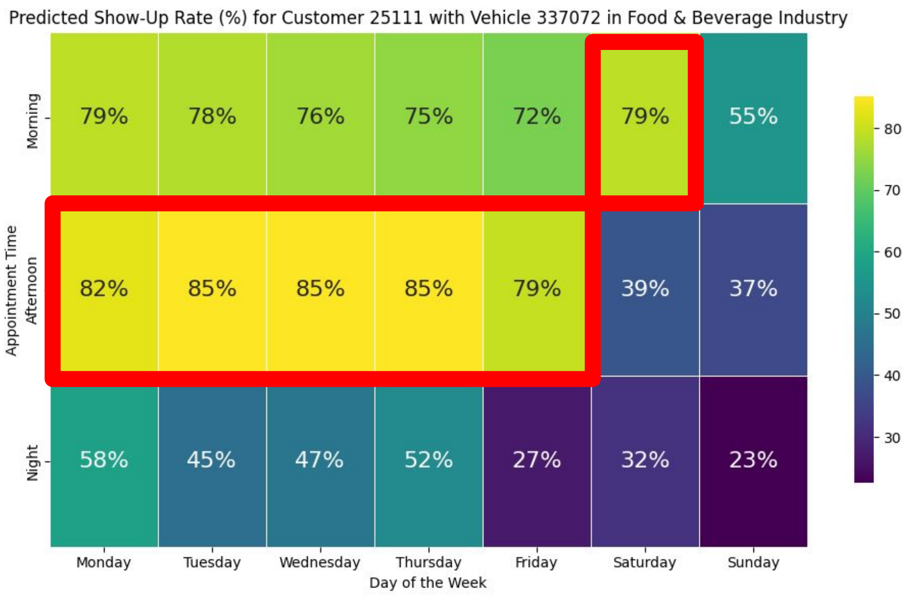
All the industries has the **Higher show-up rates on afternoons** and **Lower show-up rates on nights**.

The average show-up rate:

- Morning: 39%
- Afternoon: 48%
- Night: 18%

Specific Example for Food & Beverage Industry

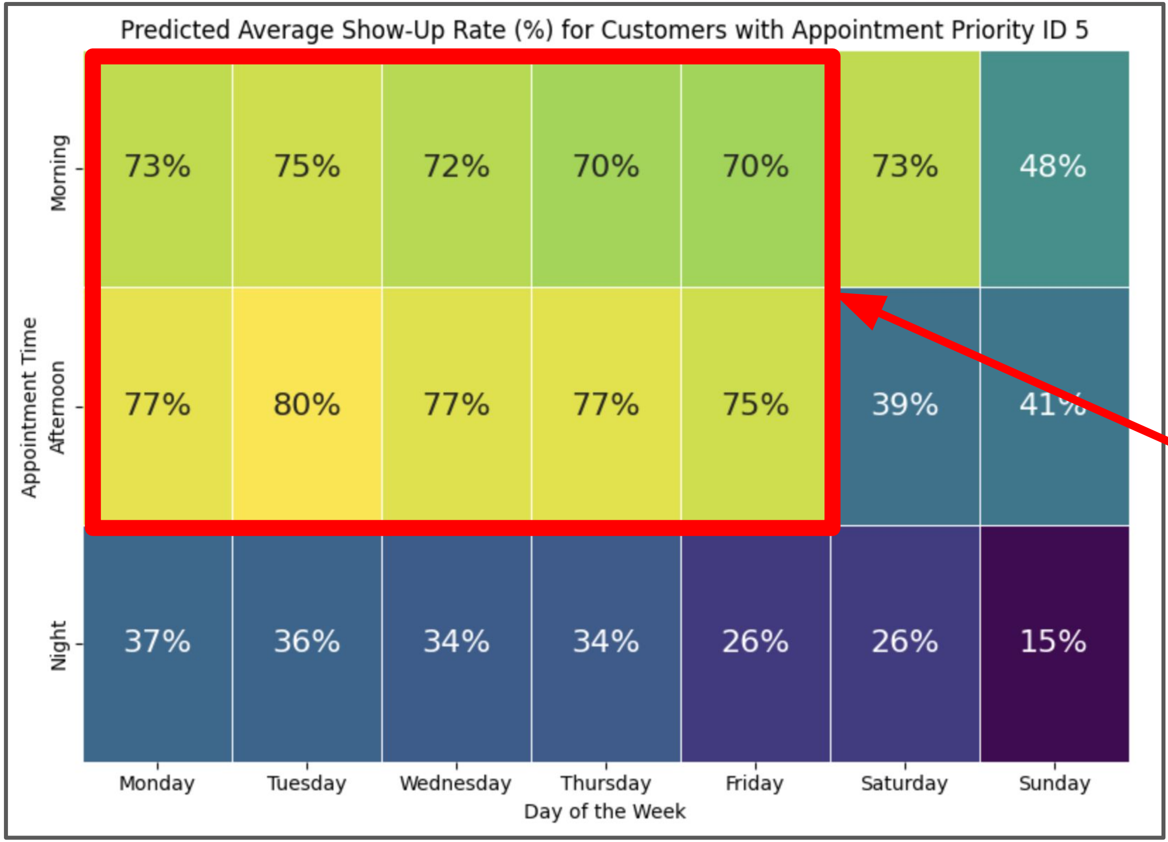




On **weekdays**, the industry is most likely to have the highest show-up rate **in the afternoon.**

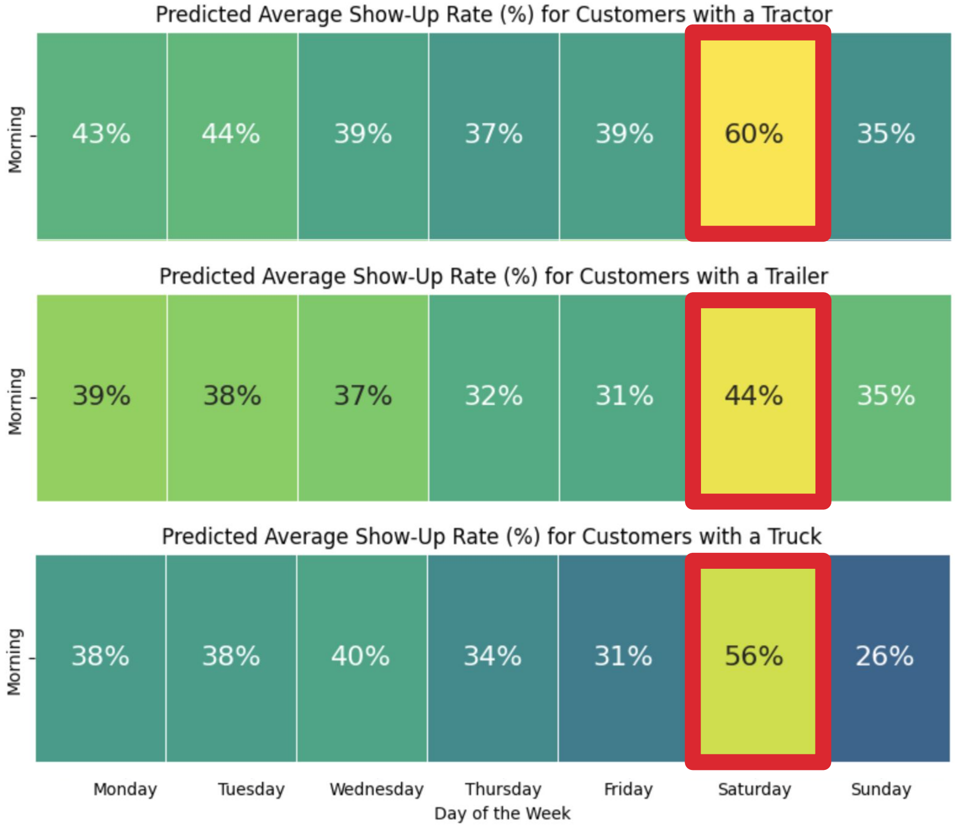
On **weekends**, the industry is most likely to have the highest show-up rate **on Saturday mornings**.

Appointment Priority ID



Customers with **appointment priority ID 5**, who have hard schedule appointments, have the highest overall probability of showing up on **weekday mornings** and **afternoons** out of all priorities.

Vehicle Type - Tractor, Trailer, and Truck

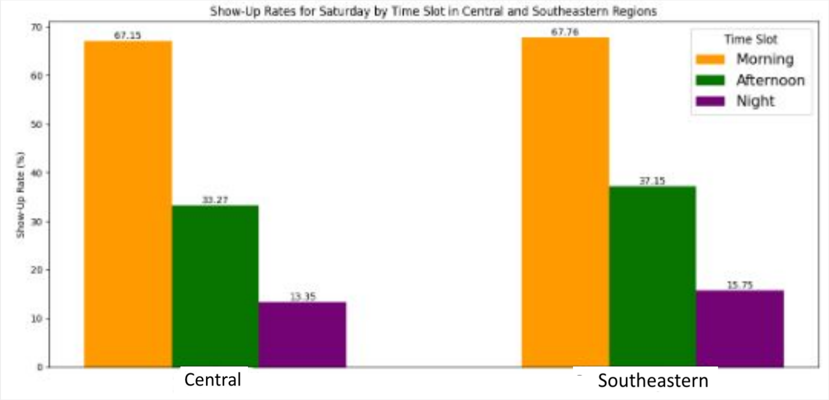


All three vehicle types show that **Saturday morning** has the highest probability of showing up out of all morning appointments.

Similar with previous insights, the complete heatmap (not pictured) of all vehicle types show higher overall likelihood of showing up on **weekday afternoons** and lower likelihood at night.

Regions

Across all regions, customers are more likely to show up in the **afternoon** and less likely to show up at night.



In Central and Southeastern regions, the highest show-up rate on Saturday is in the **morning**.

Recommendations

1. **Review Resource Allocation**
    - Review **resource allocation** to match high probability timeslots.

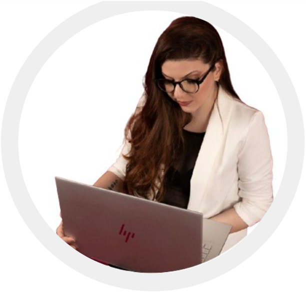
2. **Customer Engagement**
    - Conduct **surveys** and **feedback sessions** to gain insights into why customers prefer certain times or miss appointments, especially during low show-up periods.
    - Utilize personalized communication strategies (**emails**, **SMS**) to **remind customers of their appointments**, especially for those scheduled in lower probability time slots like nighttime.
    - Provide **incentives** for customers to utilize lower probability timeslots.

Special Thanks to Our Mentors!

Maria V. Gonzalez

Norman Marston

Rochak Tarika



Lenniet M. Coello

We couldn’t have done it without you ❤️

Prof. Josue Martinez

TA Elijah Hall

Q&A