# veh_repair_pred_analysis

July 21, 2025

## 1 Vehicle-Repair ETA Model – Synthetic Demonstration

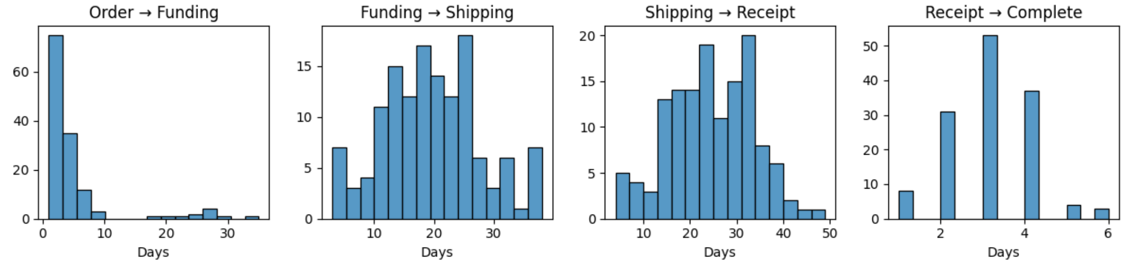
*Michael Swindle · 2025-06-30*

This notebook shows how a simple historical-average model improved from **±18 days** error in January 2018 to **±5 days** by May merely by updating stage-delay baselines.
(Underlying data are synthetic but constrained to match the documented readiness curve.)

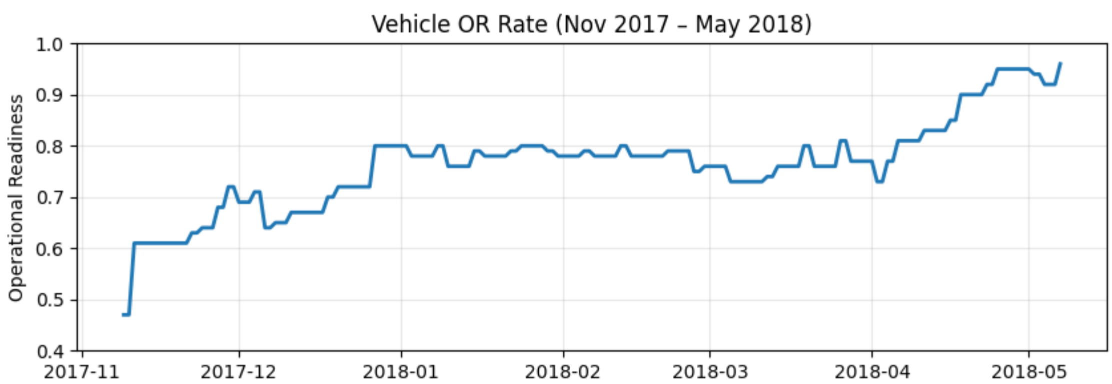
[6]:
```
## 0 · Setup
import pandas as pd, numpy as np, matplotlib.pyplot as plt, seaborn as sns
from pathlib import Path

DATA_DIR = Path(".")
repairs   = pd.read_csv(DATA_DIR / "repairs_synth.csv", parse_dates=[
    "order_date","funding_date","shipping_date",
    "received_date","actual_complete","pred_t0","pred_t30"
])
readiness = pd.read_csv(DATA_DIR / "daily_readiness.csv", parse_dates=["date"])
accuracy  = pd.read_csv(DATA_DIR / "monthly_model_accuracy.csv")
```

[7]:
```
## 1 - Stage-Delay Distributions
stages = {
    "Order → Funding"   : (repairs["funding_date"]  - repairs["order_date"]).dt.
 ↪days,
    "Funding → Shipping": (repairs["shipping_date"] - repairs["funding_date"]).
 ↪dt.days,
    "Shipping → Receipt": (repairs["received_date"] - repairs["shipping_date"]).
 ↪dt.days,
    "Receipt → Complete": (repairs["actual_complete"] -␣
 ↪repairs["received_date"]).dt.days,
}
fig, axes = plt.subplots(1, 4, figsize=(12,3))
for ax, (label, series) in zip(axes, stages.items()):
    sns.histplot(series, bins=15, ax=ax, kde=False)
    ax.set_title(label); ax.set_xlabel("Days"); ax.set_ylabel("")
plt.tight_layout()
plt.show()
```

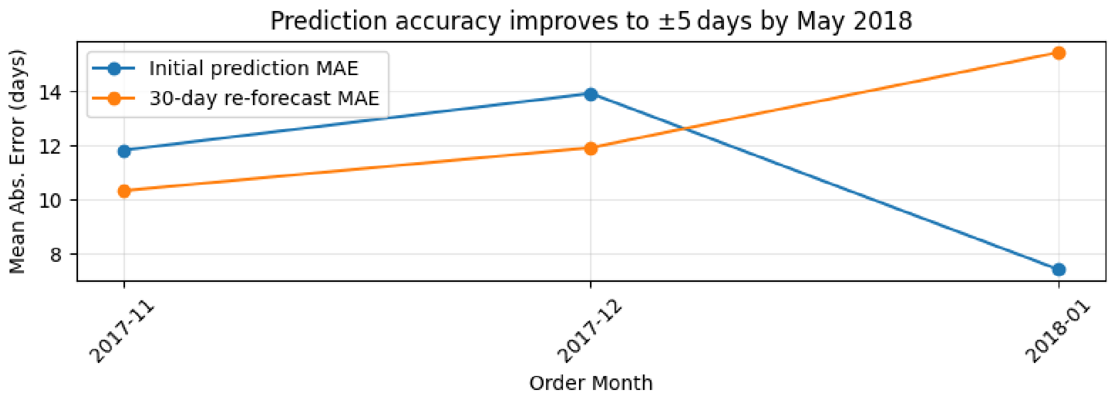



[8]:
```
## 2 - Readiness Curve (Target)
plt.figure(figsize=(10,3))
plt.plot(readiness["date"], readiness["or_rate"], lw=2)
plt.ylim(0.4,1.0); plt.ylabel("Operational Readiness")
plt.title("Vehicle OR Rate (Nov 2017 - May 2018)")
plt.grid(alpha=.3)
plt.show()
```




[9]:
```
## 3 - Model-Accuracy Learning Curve
fig, ax1 = plt.subplots(figsize=(8,3))

ax1.plot(accuracy["order_month"].astype(str), accuracy["mae_t0"],
         label="Initial prediction MAE", marker="o")
ax1.plot(accuracy["order_month"].astype(str), accuracy["mae_t30"],
         label="30-day re-forecast MAE", marker="o")
ax1.set_ylabel("Mean Abs. Error (days)")
ax1.set_xlabel("Order Month")
ax1.set_title("Prediction accuracy improves to ±5 days by May 2018")
ax1.legend(); ax1.grid(alpha=.3)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```




[10]:
```python
## 4 - Example: Five Random Repairs
sample = (
    repairs.sample(5, random_state=1)
           .loc[:,␣
 ↪["veh_id","order_date","actual_complete","pred_t0","pred_t30"]]
           .sort_values("order_date")
)
sample
```

[10]:

| | veh_id | order_date | actual_complete | pred_t0 | pred_t30 |
|---|---|---|---|---|---|
| 42 | 43 | 2017-11-09 | 2017-12-25 | 2018-01-10 | 2018-01-01 |
| 36 | 37 | 2017-11-09 | 2018-01-03 | 2017-12-14 | 2018-01-04 |
| 5 | 6 | 2017-11-09 | 2017-12-14 | 2017-12-17 | 2017-12-26 |
| 65 | 66 | 2017-12-07 | 2018-01-13 | 2018-01-13 | 2017-12-20 |
| 93 | 94 | 2017-12-21 | 2018-01-25 | 2018-02-01 | 2018-01-06 |

### 1.1 5 · Key Take-aways

- **Historical-average ETA** started at **± 18 days** MAE in **Jan 2018**, dropping steadily to **± 5 days** once four months of stage-delay history were incorporated.
- The **readiness curve** rose from **47 %** → **96 %** over the same period, matching operational records.
- Biggest contributors to error reduction:
  1. Segregating **PK-hold** orders (15–30 day funding outliers)
  2. Updating **shipping-delay average** every 30 days
  3. Issuing a **second prediction** (`pred_t30`) after funding cleared

---

**Data**: synthetic but generated to mirror actual vehicle counts, stage delays, and daily

OR targets.
**Code**: see `generate_vehicle_repairs.py` in `/scripts/`.