GP Solutions

FROM FRAGMENTED DATA TO GROWTH POWER: A BIG DATA SOLUTION FOR A EUROPEAN TOUR OPERATOR

Case Studies

-ms-user-select: none;

.button--tertiary:focus::after, .button--tertiary:hover::after {
    color: #161818
}

.button--labeled {
    padding-left: 1.143em;
    background: white;
    width: 525px;

transform: scale(1.05)
.button--tertiary:focus, .button--tertiary:hover {
    color: #fff;
    background: #fff

From Fragmented Data to Growth Power: a Big Data Solution for a European Tour Operator
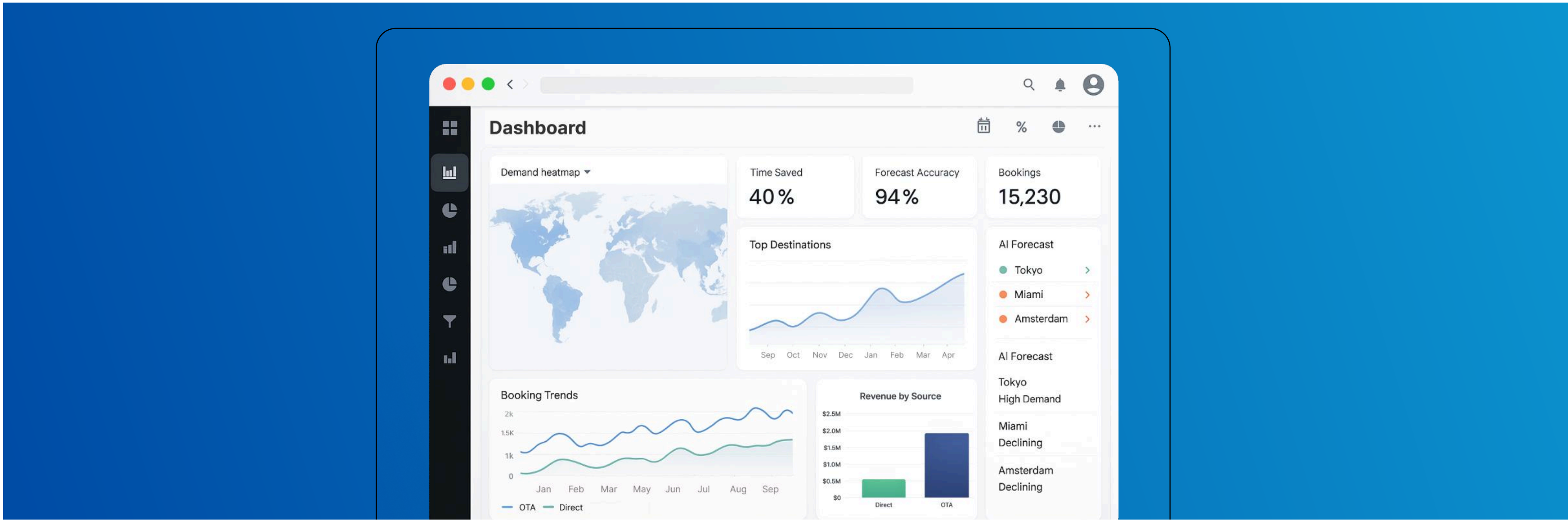
- **Tech Stack:**
AWS S3, Apache Spark, Kafka + Spark Streaming, Apache Airflow, Power BI

- **Project Management:**
Scrum

- **Start Date:**
2024

- **Scope:**
  - 6,000+ man-hours

Client Background

The client is a mid-sized European tour operator specializing in multi-destination adventure and luxury package tours across more than 25 countries. With operations spanning such diverse destinations, the company manages not only customer expectations but also a wide variety of local suppliers, seasonal demand spikes, and rapidly changing travel conditions.

They operate through a complex network of channels: their own digital booking platform, global OTAs, and traditional travel agencies with long-standing partnerships. This hybrid sales approach allows them to reach different customer demographics, but it also multiplies the complexity of data flows. Each month, thousands of itineraries, inquiries, and customer interactions are processed, creating a data environment that requires constant attention and accuracy.

However, their data was siloed across multiple disconnected systems: legacy booking software, supplier XML feeds, unstructured customer feedback forms, and external web analytics tools. These silos meant valuable insights were hidden or delayed, creating inconsistencies across reporting. This fragmentation prevented management from having a reliable "single pane of glass" view over bookings, supplier performance, and customer behavior, which in turn limited the company's ability to make confident, data-driven strategic decisions.

Client

European tour operator

Project

Big Data Platform: Data Lake, Forecasting, Dynamic Pricing, Recommendations

Key Results

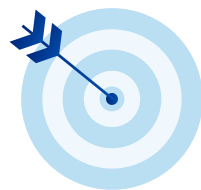| **8%**	**22%**	**15%**	**<30min**
revenue increase per booking	improved demand forecast accuracy	lift in repeat bookings	Reporting time cut

Project Goals and Challenges

The fragmented data ecosystem created more than just IT headaches—it led to significant **operational and commercial challenges** that directly affected profitability and customer satisfaction:

- **Inaccurate demand forecasting:** The operator struggled with balancing supply and demand. Popular tours were frequently overbooked during high season, leading to customer dissatisfaction and service disruptions. At the same time, less popular routes were often undersold, resulting in half-empty buses, idle guides, and wasted resources.

- **Generic marketing and low conversion:** Campaigns were largely designed with a "one-size-fits-all" approach. Without actionable segmentation, marketing teams delivered broad offers that often failed to resonate with traveler interests. This led to rising acquisition costs, missed upsell opportunities, and lower brand engagement compared to competitors offering targeted digital personalization.

- **Static, uncompetitive pricing:** Pricing decisions relied on fixed seasonal tables, which could not adapt to dynamic travel conditions. As a result, the operator frequently missed opportunities to raise margins when demand was high, and conversely, struggled to remain competitive when the market turned. Competitors leveraging dynamic pricing models were quickly gaining an edge.

- **Operational inefficiency:** Manual scheduling, route planning, and exception handling slowed down operations. Inability to respond to real-time disruptions, such as delayed flights or weather changes, meant the company lacked the agility to adjust itineraries seamlessly, diminishing overall customer experience and trust.

The primary goal of the project was to build a unified, modern data infrastructure that could consolidate all these disconnected sources, remove data silos, and serve as a foundation for real-time analytics and intelligent automation. In addition, the client wanted to enable future-ready capabilities such as predictive demand modeling, AI-driven dynamic pricing, and personalized tour recommendations.

Project and Its Development

GP Solutions partnered with the client to design and deliver a custom big data platform that **transformed fragmented data into a structured and reliable business asset.** The project aimed not only to fix current inefficiencies but also to equip the client with a foundation for long-term growth.

Using the Scrum methodology, development was carried out iteratively. Each sprint focused on building, testing, and refining a component of the solution, ensuring rapid feedback loops, transparency, and flexibility in adjusting to changing business requirements. By structuring delivery around clear sprint goals, the team was able to manage a project scope exceeding 6,000 man-hours without losing momentum.

Team Composition

2 Data Engineers:
built ingestion pipelines and the data lake.

1 Data Scientist:
developed forecasting, pricing, and recommendation models.

1 Data Architect:
defined unified data models, normalization standards, and governance.

1 DevOps Engineer:
designed scalable, reliable infrastructure.

1 Business Analyst:
coordinated stakeholder requirements.

1 Project Manager:
ensured delivery, communication, and compliance.

1. UNIFIED DATA INGESTION AND STORAGE

The foundation of the solution was an AWS S3 data lake designed to handle structured, semi-structured, and unstructured data at scale.

- **Batch ingestion:** Apache Spark jobs were configured to reliably pull supplier system data (XML/CSV), integrate with legacy booking databases, and process inputs from APIs and SFTP servers. This ensured regular synchronization of critical information such as availability and pricing.

- **Real-time streaming:** Apache Kafka combined with Spark Streaming was introduced to process live feeds from OTAs and the company website. This setup captured JSON-based booking transactions and clickstream behavior, providing an always-updated view of customer intent and sales flow.

- **Data lake architecture:** To maintain governance, the lake was divided into three zones: Staging (raw data), Curated (validated and cleaned), and Analytics (optimized for BI and ML). Anonymization protocols were applied at ingestion, ensuring that sensitive customer information was GDPR-compliant from the outset.

Once unified storage was in place, the focus shifted to advanced analytics and real-time intelligence.

- **Demand forecasting model:** Developed by the data scientist and orchestrated with Apache Airflow, the model delivered a six-month outlook with 22% greater accuracy than legacy methods. This improved planning of guides, buses, and accommodation contracts.

- **Dynamic pricing engine:** An ML-driven pricing system adjusted tour prices in real time, factoring in demand signals, competitor benchmarking, seat availability, and local events. This allowed the operator to capture higher margins in peak demand scenarios and remain competitive in slower periods.

- **Personalization engine:** Machine learning models analyzed customer history and behavior, enabling tailored recommendations. Instead of broad packages, returning customers received offers aligned with their preferences, significantly boosting engagement.

- **Operational dashboards:** Power BI was integrated with the Analytics zone, giving managers access to live dashboards covering supplier performance, booking velocity, occupancy rates, and customer sentiment. This replaced static spreadsheets and accelerated decision-making across departments.

Results and Business Impact

The implementation of the big data platform by GP Solutions delivered measurable and transformative results. Within months, the client's ability to forecast, price, and market effectively improved, and operational teams reported unprecedented agility.

- **8% increase in average booking revenue:** The dynamic pricing engine ensured maximum yield per tour, directly boosting profitability.

- **22% improvement in forecast accuracy:** More reliable forecasts minimized both overbookings and underutilized resources, allowing better contract negotiations with suppliers.

- **15% lift in repeat customer bookings:** Personalized offers encouraged loyalty, driving repeat business and strengthening long-term customer relationships.

- **95% reduction in reporting time:** Reports that previously took two days to prepare were now generated in under 30 minutes, freeing management capacity and enabling data-driven agility.

- **Enhanced supplier negotiations:** Clear data insights empowered procurement teams to secure better terms, strengthen partnerships, and enforce quality standards.

This new data-driven foundation positioned the client to operate more profitably, expand their market reach, and provide superior travel experiences. By consolidating fragmented data into a unified system, the operator gained the competitive edge necessary to thrive in Europe's evolving travel industry.

GP Solutions
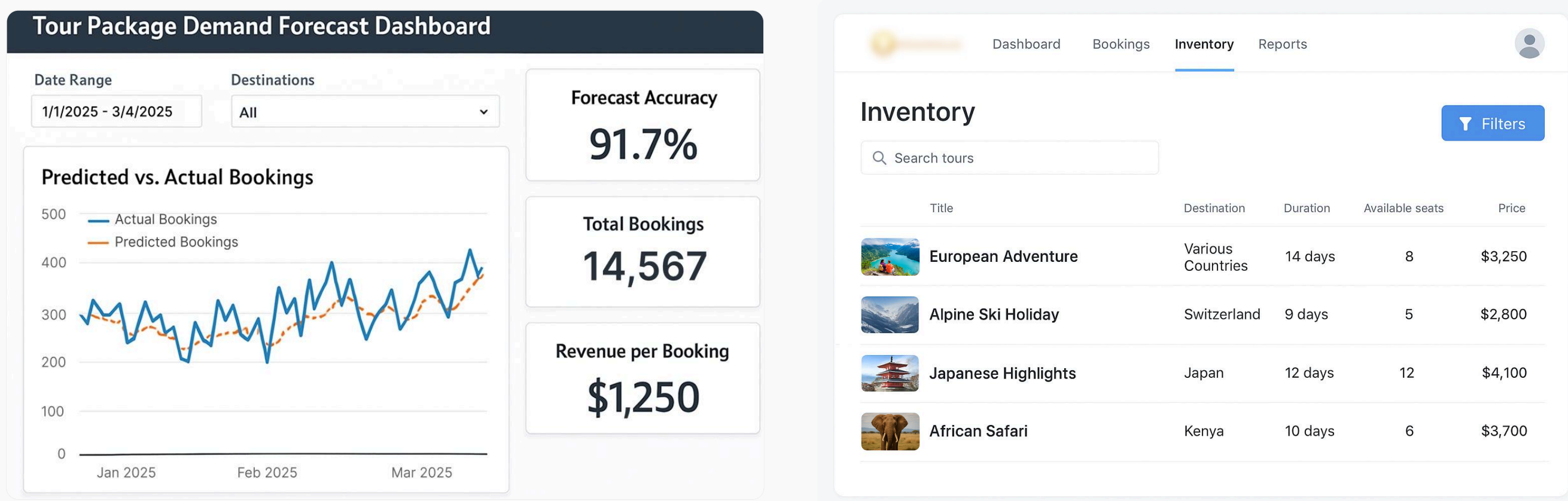
Let's Build the Future of Travel, Together.

Schedule a Strategic Consultation

✉ **Letstalk@Software.Travel**

Offices worldwide. Service everywhere. • USA • Germany • UAE • Poland