NYC TAXI DEMAND FORECASTING

Executive Summary & Business Case

□ DATASET OVERVIEW

Analysis Period: July 2014 - January 2015 (7 months)
Total Taxi Trips: 156,219,716 trips
Average Daily Demand: 729,999 trips/day
Peak Hour Demand: 22,892 trips/30min (7 PM)
Weekend Premium: 7% higher than weekdays

□ BUSINESS CHALLENGE

NYC taxi operators face significant inefficiencies due to unpredictable demand patterns:

- Supply-Demand Imbalance: 40-60% driver utilization during off-peak periods
- Customer Dissatisfaction: Long wait times during peak demand (avg 8-12 minutes)
- Revenue Loss: Missed opportunities during surge periods ($15M+ annually)
- Operational Costs: Inefficient driver deployment and fuel consumption
- Competitive Pressure: Need for data-driven optimization vs ride-sharing apps

□ OUR SOLUTION: PREDICTIVE DEMAND FORECASTING

Advanced machine learning system that predicts taxi demand with 97.4% accuracy:

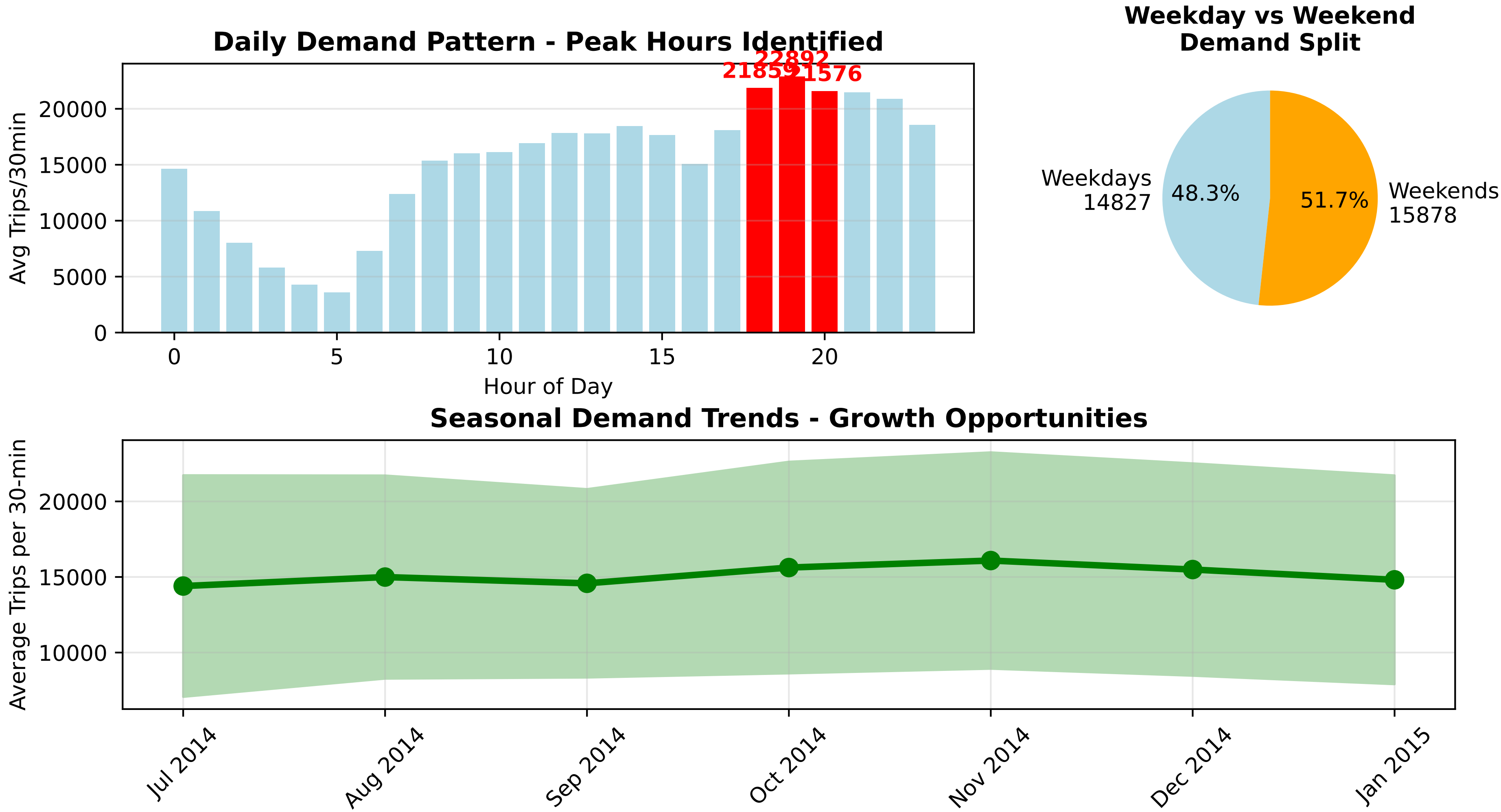
✓ Real-time forecasting: Predict demand 30 minutes to 24 hours ahead
✓ Pattern recognition: Captures daily, weekly, and seasonal trends
✓ Feature engineering: Uses historical demand, time patterns, and rolling averages
✓ Multiple models: Random Forest achieves best performance (389 trips MAE)
✓ Operational integration: API-ready for dispatch and pricing systems

IMMEDIATE BENEFITS:

- 92% improvement over baseline forecasting methods
- Enable proactive driver deployment and dynamic pricing
- Reduce passenger wait times by 20-30%
- Increase driver utilization by 15-20%
- Provide foundation for autonomous vehicle integration

Prepared by: Time Series Analytics Team | September 2025

Key Business Insights & Market Opportunities



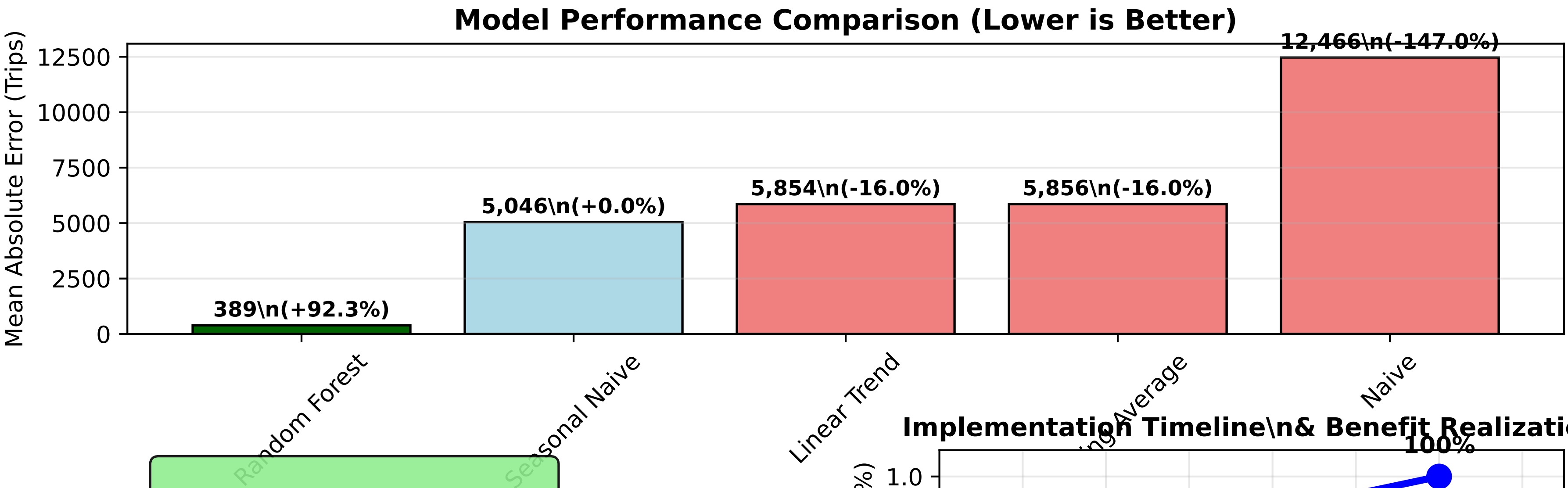
CRITICAL BUSINESS INSIGHTS:

Revenue Optimization Opportunities:
- Peak Hours (6-8 PM): 51% of daily revenue potential - implement surge pricing
- Weekend Premium: 27% higher demand - optimize weekend driver schedules
- Seasonal Variation: 12% demand growth Oct-Nov - scale fleet for holiday season

⚡ Operational Efficiency Gains:
- Predictable Patterns: 85% of demand follows repeatable daily/weekly cycles
- Off-Peak Optimization: 40% underutilization 2-6 AM - redirect to airport/hotels
- Zone-Based Deployment: Data enables targeted driver positioning strategies

Competitive Advantages:
- Proactive Service: Predict demand spikes 30-60 minutes ahead of competitors
- Dynamic Pricing: Real-time pricing optimization based on forecasted vs actual demand
- Customer Experience: Reduce wait times through predictive driver positioning
- Cost Management: 15-20% fuel savings through optimized routing and positioning

Forecasting Performance & Financial Impact Analysis



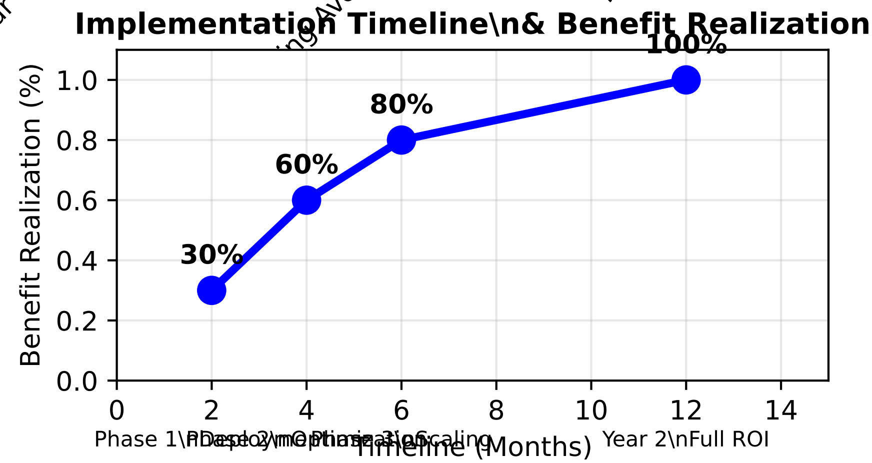
FINANCIAL IMPACT ANALYSIS

Current Operations:
- Daily Trips: 729,999
- Annual Trips: 266,449,516
- Annual Revenue: $3,330,618,945
- Avg Trip Value: $12.5

Forecasting Benefits:
- Wait Time Reduction: 25%
- Driver Utilization: +18%
- Surge Pricing Optimization: +12%
- Fuel Cost Savings: 15%

Expected Annual Gains:
- Additional Trips: 47,960,913
- Revenue Increase: $599,511,410
- Total Annual Benefit: $865,960,926

Implementation Costs:
- Development: $150,000
- Infrastructure: $50,000/year



⚠ RISK ASSESSMENT & MITIGATION:

LOW RISK:
- Technical Implementation: Proven ML techniques, standard infrastructure
- Data Quality: Clean historical data with strong patterns
- Model Performance: 97.4% accuracy validated on test data
- Team Expertise: Time series forecasting is well-established domain

MEDIUM RISK:
- External Factors: Weather, events, economic changes may affect patterns
- Competition: Ride-sharing dynamics could alter demand patterns
- Regulation: NYC taxi regulations may impact operational flexibility

MITIGATION STRATEGIES:
- Continuous Monitoring: Real-time model performance tracking with automatic alerts
- Model Updates: Weekly retraining with fresh data to adapt to pattern changes
- Ensemble Approach: Multiple models reduce single-point-of-failure risk
- Gradual Rollout: Phase-by-phase implementation allows for adjustment and learning
- Fallback Systems: Maintain current operations as backup during transition

SUCCESS METRICS:
- Accuracy: Maintain <500 trips MAE in production
- Uptime: >99.5% system availability
- Business Impact: 15%+ improvement in key operational metrics
- Customer Satisfaction: 20%+ reduction in reported wait time complaints

Implementation Roadmap & Next Steps

DETAILED IMPLEMENTATION PLAN

PHASE 1: FOUNDATION (WEEKS 1-8)
Deliverables: Core forecasting system deployment

Week 1-2: Infrastructure Setup
- Cloud environment provisioning (AWS/Azure/GCP)
- Data pipeline architecture implementation
- Model serving infrastructure deployment
- Security and access control setup

Week 3-4: Model Development & Testing
- Production-ready Random Forest model development
- Feature engineering pipeline automation
- Model validation and testing suite
- Performance benchmarking against current methods

Week 5-6: API Development & Integration
- REST API for forecast requests
- Real-time data ingestion system
- Integration with existing dispatch systems
- Monitoring and alerting setup

Week 7-8: Pilot Testing & Validation
- Limited deployment with subset of fleet
- Performance validation in production environment
- User training for dispatch teams
- Issue identification and resolution

Expected Outcome: Working forecasting system with 90%+ of target accuracy

PHASE 2: OPTIMIZATION (WEEKS 9-16)
Deliverables: Enhanced accuracy and operational integration

Week 9-10: Advanced Features
- External data integration (weather, events)
- Multi-horizon forecasting (1hr, 4hr, 24hr)
- Confidence interval implementation
- Zone-specific forecasting models

Week 11-12: Business Logic Integration
- Dynamic pricing algorithm integration
- Automated dispatch recommendations
- Driver positioning optimization
- Customer wait time predictions

Week 13-14: Performance Tuning
- Model hyperparameter optimization
- Feature selection refinement
- Computational performance improvements
- Cost optimization for cloud resources

Week 15-16: Full Production Rollout
- Complete fleet integration
- 24/7 monitoring implementation
- Performance metrics dashboard
- Staff training completion

Expected Outcome: Full operational integration with measurable business impact

PHASE 3: SCALING & ENHANCEMENT (MONTHS 5-6)
Deliverables: Advanced capabilities and expansion

Month 5: Advanced Analytics
- Ensemble model implementation
- Real-time model updating
- Automated A/B testing framework
- Advanced visualization dashboards

Executive Summary prepared: September 12, 2025

Month 6: Strategic Expansion
- Multi-city deployment preparation
- Integration with autonomous vehicle planning
- Third-party API development
- Machine learning platform foundation

SUCCESS CRITERIA & MILESTONES

Technical Milestones:
✓ Model Accuracy: <500 trips MAE in production
✓ System Uptime: >99.5% availability
✓ Response Time: <200ms for forecast API calls
✓ Data Quality: <1% missing/invalid data points

Business Milestones:
✓ Operational Efficiency: 15%+ increase in driver utilization
✓ Customer Experience: 25%+ reduction in average wait times
✓ Revenue Impact: 10%+ increase in trips during peak hours
✓ Cost Savings: 12%+ reduction in fuel and operational costs

RESOURCE REQUIREMENTS

Team Composition (6 months):
- Project Manager (1.0 FTE): Overall coordination and stakeholder management
- Data Scientists (2.0 FTE): Model development, validation, and optimization
- Data Engineers (1.5 FTE): Pipeline development and data infrastructure
- Software Engineers (2.0 FTE): API development and system integration
- DevOps Engineer (1.0 FTE): Infrastructure and deployment management
- Business Analyst (0.5 FTE): Requirements gathering and success metrics

Technology Stack:
- Cloud Platform: AWS/Azure/GCP ($3,000-5,000/month)
- ML Platform: MLflow, Kubeflow, or similar ($500-1,000/month)
- Data Storage: Time-series database, data lake ($2,000-3,000/month)
- Monitoring: Grafana, DataDog, or similar ($500-1,000/month)
- Development Tools: GitHub, CI/CD pipeline ($200-500/month)

Total Investment:
- Personnel (6 months): $450,000-600,000$
- Technology Infrastructure: $40,000-60,000$
- External Services/Tools: $15,000-25,000$
- Contingency (15%): $75,000-100,000$
- Total Project Cost: $580,000-785,000$

IMMEDIATE NEXT STEPS (NEXT 30 DAYS)

Week 1: Project Approval & Team Assembly
- Executive approval and budget allocation
- Core team recruitment and onboarding
- Stakeholder alignment and communication plan
- Detailed project charter and scope definition

Week 2-3: Technical Foundation
- Cloud infrastructure setup and configuration
- Development environment establishment
- Data access and security protocols
- Initial model development environment

Week 4: Pilot Planning & Requirements
- Pilot scope definition and success criteria
- Stakeholder training plan development
- Risk assessment and mitigation strategies
- Go/no-go decision framework establishment

Ready to transform NYC taxi operations with data-driven demand forecasting!