Scenario-Based Neural Network Planning for Urban Public Transport Demand Under Digital Transformation: Evidence from Tashkent

Vazira Nazarova[1], Barno Kenjayeva[1] and Saule Ilasheva[2]
[1]*Department of Transport Logistics, Tashkent State Transport University, Temiryolchilar Str. 1, 100167 Tashkent, Uzbekistan*
[2]*Department of Economics, M. Auezov South Kazakhstan University, Tauke Khan Ave. 5, 160012 Shymkent, Kazakhstan*
nvazira68@gmail.com, barnokenjayeva4@gmail.com, saule.Ilasheva@auezov.edu.kz

Keywords: Passenger Demand, Public Transport Planning, Scenario Analysis, MLP Regressor, Conditional Hold-Out Check, Sustainable Urban Mobility.

Abstract: In the context of rapid urbanization in Tashkent and ongoing Smart City reforms, deterministic transport-planning approaches may inadequately capture non-linear shifts in recorded passenger demand under structural change. This study proposes a planning-oriented scenario framework combining a multilayer perceptron (MLP) regressor with an externally specified logistic guardrail to support public transport capacity analysis through 2030. The empirical basis comprises anonymized annual operational reports from eight bus depots for 2014-2023 and rolling-stock inventory data as of 01.06.2025. A linear pre-shock trend is retained only as a conservative baseline illustrating the limits of simple extrapolation after disruption. Given the very small annual sample and the single hold-out check for 2023, the framework is not presented as evidence of stable forecasting performance. Instead, it serves as a decision-support instrument mapping exogenously specified digital trip-capture assumptions into conditional demand trajectories under a conservative normative upper bound and linking them to rolling-stock planning implications for long-horizon public transport planning needs.

1 INTRODUCTION

The sustainable development of megacities in the Industry 4.0 paradigm [1] depends on balancing mobility demand with the efficient use of public transport capacity [2]. In Tashkent, where rapid demographic growth [3] coincides with accelerating motorization, improving public transport performance has evolved from a local logistics issue into a macroeconomic task [4].

A major source of pressure on transport infrastructure is the spatial transformation of Tashkent. A construction boom, including vertical residential densification and the incorporation of adjacent districts into the city boundary, has intensified travel demand [5]. The opening of new universities and business centers has increased labor- and education-related mobility; although the official resident population is 3.1 million, the effective daily population rises by 30-35% or more. Over the past eight years, the number of vehicles has doubled. At the same time, the expansion of digital fare-system reforms and related digital trip registration appears to have increased the statistical visibility of previously under-recorded passenger activity [6]. Together, these processes have reshaped established passenger flows [7].

Under these conditions, transport-operator competitiveness depends on the adaptive management of carrying capacity. Retrospective operating data from Tashkent bus depots, including internal reports and technical-operational indicators for 2014-2023 from Toshshahartransxizmat JSC [8], reveal the limitations of deterministic planning methods [9]. Linear extrapolation of historical trends is adequate only under stationary conditions.

Statistics for 2020-2023 illustrate the fragility of inertial planning under structural shocks. Passenger flows fell to 104.7 million in 2020, while the integration of bank cards into public-transport payment [10] coincided with a 67.3% rise in electronically registered passenger trips in 2023. This increase reflects greater statistical visibility of previously under-recorded travel rather than a single causal mechanism. Against this background, the study develops and applies an adaptive neural-network-based scenario framework for policy-relevant public transport planning under market recovery, formalization-related changes in recorded statistics, and transport-system reforms [11].

The contributions of this paper are as follows:

1) Compilation of a structured annual dataset from anonymized operational reports of eight Tashkent bus depots for 2014-2023, supplemented with rolling-stock inventory data.
2) Implementation of a leakage-safe conditional hold-out design separating the training period (2014-2022) from the previously unseen year 2023.
3) Development of a hybrid scenario-based planning framework in which an MLP captures non-linear conditional response patterns, while a normative logistic guardrail limits long-horizon trajectories to a conservative realism corridor.
4) Use of exogenously specified $Y_{digital}$ inputs as policy-conditioned assumptions on digital trip capture in recorded statistics rather than as endogenous forecasts.
5) Translation of conditional demand trajectories into rolling-stock planning implications under alternative operational-efficiency assumptions through 2030.

It is important to emphasize that the reported metrics (in-sample MAPE = 4.3% for 2020-2022 and hold-out APE = 1.5% for 2023) refer only to this single conditional hold-out year. Because the test set consists of one observation, these figures constitute a restricted case-specific check and do not demonstrate stable generalization capability. All subsequent claims regarding model performance are made with this limitation in mind.

2 MATERIALS AND METHODS

2.1 Dataset Construction and Preprocessing

The empirical basis consists of official anonymized operational reports of Toshshahartransxizmat JSC across eight Tashkent bus depots (No. 1, 2, 4, 5, 7, 8, 12, 18). The annual passenger-demand series (Y_{demand}) and the observed series of electronically registered trips in recorded statistics ($Y_{digital}$) were taken directly from operational reports for 2014-2023. Interpolation was applied only to selected auxiliary operational indicators for 2014-2017 where the consolidated archive lacked complete entries; the target demand variable was not interpolated. Lagged fleet size $N_{bus(t-1)}$ was included as a predictor to mitigate simultaneity between current fleet size and current demand. For 2019-2023, lagged fleet-size values were taken directly from annual depot summaries. For 2014-2018, the corresponding lagged entries were unavailable in the consolidated archive and were retrospectively reconstructed as author estimates by backward linear extrapolation from adjacent annual fleet information. These early-year values were used only as auxiliary predictor inputs to maintain a consistent annual feature matrix for 2014-2022. Feature scaling was based only on training-sample statistics to prevent leakage. For the 2023 retrospective test, actual $Y_{digital}$ values were supplied as observed inputs; for the 2030 forecast, this variable was specified exogenously via scenarios. The 2023 result should therefore be interpreted as a single conditional hold-out check rather than as an unconditional ex ante forecast. To formalize the feature space of the neural network, the following vector of variables was selected (Table 1).

Table 2 presents the annual values of the model variables used both for model training (2014-2022) and for the single-year hold-out test (2023). For X_t, Y_{demand}, $Y_{digital}$, and K_{tg}, all entries are observed. For $N_{bus(t-1)}$, values for 2019-2023 are observed, whereas values for 2014-2018 were retrospectively reconstructed as author estimates by backward linear extrapolation and used only as auxiliary predictor inputs in model training.

No interpolation was applied to Y_{demand} or $Y_{digital}$; the only reconstructed inputs were early-year lagged fleet values obtained by backward linear extrapolation. From an interpretive perspective, X_t captures secular time and post-shock timing, $Y_{digital}$ reflects the degree of digital trip capture in official statistics, K_{tg} represents the technical supply constraint, and lagged $N_{bus(t-1)}$ approximates carrying capacity while mitigating simultaneity between current fleet size and current demand. The stronger post-pandemic co-movement between Y_{demand} and $Y_{digital}$ is consistent with broader digital capture in the accounting system, although it does not prove a single causal mechanism.

Table 1: Structure of study variables.

Notation	Variable	Unit	Description and role in the model
X_t	Time period	Year (normalized)	Predictor. Normalized relative to the start of observation (t - 2014)
Y_{demand}	Transport volume (passenger flow)	thousand passenger trips	Target variable. Indicator of realized demand
$Y_{digital}$	Electronically registered passenger trips	thousand passenger trips	Indicator of digital trip capture in recorded statistics
K_{tg}	Technical availability coefficient	0 to 1	Supply-side limiting factor
N_{bus}	Fleet size (listed stock)	units	Predictor. Lagged fleet-size input; in scenario calculations, treated as a controllable capacity parameter.

Note: Table compiled by the author.

Table 2: Annual values of model variables used in training (2014-2022) and single-year hold-out testing (2023).

Indicator	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
X_t (normalized)	0	1	2	3	4	5	6	7	8	9
Y_{demand} (thousand passenger trips)	235 903	220 165	197 785	205 784	211 260	220 577	104 727	138 209	144 967	240 403
$Y_{digital}$ (thousand passenger trips)	107 137	100 405	89 468	98 238	111 179	116 650	58 940	103 485	108 243	195 199
K_{tg}	0.785	0.760	0.745	0.730	0.707	0.721	0.593	0.708	0.633	0.795
$N_{bus(t-1)}$ (units)	1413.4	1384.5	1355.7	1326.8	1297.9	1 312.2	1 308.5	1 028.9	1 169.3	1 237.4

Note: X_t, Y_{demand}, $Y_{digital}$, and Kt_g are observed annual values taken directly from anonymized operational reports of Toshshahartransxizmat JSC. For $N_{bus(t-1)}$, values for 2019-2023 are observed, whereas values for 2014-2018 are retrospectively reconstructed author estimates obtained by backward linear extrapolation.

2.2 Mathematical Modeling

A comparative analysis of two mathematical paradigms was conducted. Linear trend extrapolation for 2014-2019 was used as a conservative pre-shock baseline representing inertial planning before the structural break, rather than as a symmetric competitor to the MLP trained on 2014-2022 data.

Analytical form:

$$Y_{lin}(t) = \alpha \cdot (t - t_0) + \beta \quad (1)$$

where:

- $Y_{lin}(t)$ is the demand predicted by the linear benchmark at time $\boldsymbol{t}$;
- α is the slope coefficient;
- β is the intercept;
- t is the calendar year;
- $t_0 = 2014$ is the reference year used for time normalization.

Time normalization improves interpretability of the intercept and makes the benchmark specification more transparent. Eq. (1) is therefore interpreted only as a conservative benchmark representing inertial planning.

The neural-network model is based on a multilayer perceptron architecture [12]. Its flexibility makes it suitable for approximating complex non-linear relationships, while prior research has shown that optimization regime and training dynamics are important determinants of neural-network generalization [13]. This is particularly important for modeling the V-shaped recovery after COVID-19 and the step-like growth associated with demand formalization in recorded statistics.

Although the annual sample is very small and does not support strong claims about model identification or stable generalization, an MLP regressor was retained here only as a flexible functional component within a planning-oriented scenario framework. Its role is not to establish predictive superiority over simpler forecasting methods, but to provide a conditional non-linear response surface under structurally disrupted annual dynamics, including the post-pandemic recovery and the step-like increase in digitally captured trips. For this reason, all subsequent results are interpreted conservatively as scenario-conditioned planning outputs rather than as evidence of robust autonomous forecasting performance.

Given the final MLP specification hidden_layer_sizes = (100, 50), forward propagation is represented by two hidden-layer transformations and one output mapping. The first hidden layer is defined by Eq. (2), the second hidden layer by Eq. (3), and the output layer by (4):

$$h_{1,j} = ReLU\left(\sum_{i=1}^{n} w_{ji}^{(1)} x_i + b_j^{(1)}\right), \quad (2)$$

$$h_{2,m} = ReLU\left(\sum_{j=1}^{p} w_{mj}^{(2)} h_{1,j} + b_m^{(2)}\right), \quad (3)$$

$$y_{pred} = \sum_{m=1}^{q} w_m^{(3)} h_{2,m} + b^{(3)}, \quad (4)$$

where x_i is the i-th input feature; $h_{1,j}$ and $h_{2,m}$ are the activations of neurons j and m in the first and second hidden layers, respectively; n is the number of input features; $p = 100$ and $q = 50$ are the numbers of neurons in the first and second hidden layers; $w_{ji}^{(1)}$, $w_{mj}^{(2)}$, and $w_m^{(3)}$ are the corresponding connection weights; $b_j^{(1)}$, $b_m^{(2)}$, and $b^{(3)}$ are the corresponding bias terms.

Both hidden layers use the rectified linear unit activation function, defined as $\mathrm{ReLU}(z) = \max(0, z)$, which enables the model to capture non-linear trend breaks while reducing the risk of vanishing gradients.

Training was performed using the mean squared error (MSE) loss function and the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) optimizer, which is well suited to small samples.

Under the available data conditions, the MLP is not claimed to be statistically superior to simpler forecasting methods in a general sense. It was retained in this study because the observed annual dynamics include a structural break, post-pandemic recovery, and a step-like increase in digitally captured trips, for which a purely linear specification is substantively restrictive. The linear benchmark was therefore used as a conservative inertial reference, while the MLP was used only as a flexible conditional mapping component within a scenario-support framework. Accordingly, the purpose of the MLP in this study is not benchmark dominance, but flexible conditional representation under structural disruption.

2.3 Experimental Implementation

The computational experiment was implemented in Python 3.9 using Scikit-learn, Pandas, and Matplotlib, consistent with common practice in transport modeling [14]-[18].

Normalization. Input variables were mapped to the [0, 1] range using MinMax scaling. As shown in (5), the scaling parameters were estimated on the training subset only in order to avoid leakage from future observations:

$$x'_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (5)$$

where x'_{scaled} is the normalized value of feature x, and x_{min} and x_{max} are the minimum and maximum values of that feature in the training sample.

Hyperparameters. The final MLP specification used in the computational experiment was hidden_layer_sizes = (100, 50), activation = 'relu', solver = 'lbfgs', alpha = 0.0001, max_iter = 5000, and random_state = 42. Accordingly, in Eqs. (2)-(4), $p = 100$ and $q = 50$. These settings were fixed before the final hold-out evaluation for 2023 in order to preserve a stable small-sample specification rather than to maximize fit for a single test year. No hyperparameter optimization was performed against the 2023 hold-out year; all specification choices were restricted to training-period considerations only.

For reproducibility, the pipeline was as follows: 1) compilation of the annual dataset for 2014-2023 from anonymized depot reports; 2) division into T_{train} = 2014-2022 and T_{test} = 2023; 3) MinMax scaling fitted on T_{train} and applied to T_{test} with the same parameters; 4) training of the MLPRegressor with the fixed hyperparameter configuration reported above; 5) generation of the 2023 hold-out forecast and evaluation by absolute percentage error (APE), numerically equivalent to MAPE for a single observation; 6) specification of exogenous $Y_{digital}$ inputs for 2024-2030 as a monotonic saturation scenario anchored at the observed 2023 value, with larger annual increments in 2024-2027 and smaller ones in 2028-2030, followed by generation of the final trajectory as the pointwise minimum of the unconstrained MLP output and the auxiliary logistic guardrail approaching the conservative ceiling of 650 million passenger trips per year. The structure of the analytical variables, the main data origins, the backward linear extrapolation rule used for early-year $N_{bus(t-1)}$ inputs, and the imposed $Y_{digital}$ scenario values for 2024-2030 are reported directly in the manuscript.

2.4 Conditional Hold-Out Check and Scenario Methodology

To ensure an objective accuracy assessment, an out-of-sample evaluation protocol was applied, consistent with standard testing of predictive models on data excluded from model construction and optimization:

1) The model is trained on the Ttrain horizon through 2022, inclusive.
2) A forecast is generated for 2023.

3) The forecast is compared with the observed 2023 value, and absolute percentage error (APE) is calculated.

Scenario-based demand planning through 2030 was conducted within a normative-target framework. Boundary conditions for modeling (saturation constraints) were derived from the demographic forecast for the Tashkent agglomeration population by 2030 (approximately 3.2-3.6 million people) [19] and a target public-transport mobility coefficient consistent with sustainable transport-system standards (200-210 trips per resident per year) [20]. Based on these assumptions, the saturation ceiling was fixed at 650 million passenger trips per year. This value should be interpreted as a conservative normative upper bound for long-horizon planning rather than as an empirically observed level. Numerically, the lower boundary of this corridor equals 3.2 x 200 = 640 million trips, whereas the upper boundary equals 3.6 x 210 = 756 million trips; the adopted ceiling of 650 million therefore lies close to the conservative lower bound.

To operationalize the saturation mechanism, an auxiliary logistic guardrail was introduced as a normative upper-bound path for long-horizon planning rather than as an independently estimated empirical forecast. The guardrail approaches the conservative ceiling K = 650,000 thousand passenger trips per year and was anchored at the observed 2023 demand level.

$$Q_t^{log} = \frac{K}{1+A\, e^{-r(t-2023)}}, \quad (6)$$

where Q_t^{log} is the logistic guardrail in year t, $Q_{2023}^{obs} =$ 240,403 thousand passenger trips is the observed demand in 2023, and $A = \frac{K}{Q_{2023}^{obs}} - 1$ ensures anchoring at the starting point of the forecast horizon. The growth-rate parameter r was calibrated so that the auxiliary guardrail approached a scenario-consistent terminal level of $Q_{2030}^{log} =$ 595,800 thousand passenger trips by 2030, corresponding to 91.7% of the conservative ceiling, which yields r of approximately 0.419 per year. This terminal level was used only to parameterize the external guardrail and did not serve as a fitting target for the MLP. Accordingly, the logistic component was specified externally as a normative realism constraint rather than fitted as a separate data-driven model on the small annual sample.

$$\hat{Q}_t = min\left(\hat{Q}_t^{MLP}, Q_t^{log}\right), t = 2024, \ldots, 2030, \quad (7)$$

The logistic guardrail (Eq. 6) is not a data-driven forecast but an externally calibrated normative upper bound derived from official demographic projections and sustainable mobility targets. It serves only as a realism constraint and was not fitted to the training data. The saturation mechanism was implemented in two steps: first, the MLP generated an unconstrained trajectory conditional on exogenous inputs; second, this trajectory was compared with the logistic guardrail, and the lower value was retained as the final planning forecast. The 2024-2030 trajectory should not be interpreted as an endogenous model prediction. $Y_{digital}$ was imposed exogenously, while the logistic guardrail was specified as an external normative constraint. Under the present calibration, the guardrail remained above the unconstrained MLP path for most of the horizon and became relevant only near 2030. The terminal value of 595.8 million passenger trips was used only to parameterize the external guardrail and did not serve to fit or tune the MLP, which was estimated exclusively on the 2014-2022 training sample under the fixed specification.

3 RESULTS

3.1 Conditional Retrospective Check for the Previously Unseen Year 2023

This section reports a limited conditional retrospective check rather than full validation of forecasting performance. Model behavior was examined separately for in-sample approximation over the training period and for the previously unseen year 2023 under observed covariates. Given the very small annual sample and the single hold-out observation, this exercise serves only to assess whether the fitted conditional relationships remain planning-relevant beyond the estimation period.

Results are reported in Table 3, where training-period approximation and test-period forecasting are separated for clarity. For the linear benchmark, values for 2020-2023 are extrapolations from the 2014-2019 trend and should not be read as training-period approximation.

Table 3 should be read primarily as an illustration of conditional planning relevance under structural disruption rather than as evidence of superior forecasting performance. The linear benchmark, estimated on the pre-shock trend, performs poorly after the break, whereas the MLP yields a closer conditional approximation for the single previously unseen year 2023 under observed 2023 inputs. Because the out-of-sample assessment is based on a single year and the information set differs across methods, this comparison should be interpreted only as a limited scenario-support check rather than as evidence of robust predictive superiority. For compactness, the main text reports only the post-shock training subperiod 2020-2022, which is substantively central to the structural-break argument.

Visual confirmation is provided in Figure 1: the neural-network model (blue dotted line) closely tracks the V-shaped recovery, whereas the linear benchmark (red dashed line) remains comparatively inertial.

In Figure 1, the MLP line shows training-period approximation for 2020-2022 and a single conditional hold-out check for 2023 under observed covariates, whereas the linear line shows inertial post-shock extrapolation from the 2014-2019 baseline.

3.2 Structural Transformation of Demand

A key feature of the 2023 increase appears to be not only physical growth in mobility but also the formalization of passenger flows in the recorded statistics. A comparison of the pre-pandemic year 2019 and the reporting year 2023 in Table 4 illustrates this interpretation. The changes reported in Table 4 should be interpreted descriptively, as shifts in recorded operating indicators, rather than as direct identification of a single causal mechanism.

The descriptive contrast is also visible in the composition of recorded statistics: the ratio $Y_{digital}$ / Y_{demand} increased from 52.9% in 2019 to 81.2% in 2023. This shift does not by itself prove a single causal mechanism. It is consistent with the descriptive interpretation of increased statistical visibility of passenger trips under digital registration. No single causal mechanism is claimed or identified.

Table 3: Separation of training-period approximation and test-period forecast results for the linear benchmark and MLP.

Year	Actual, thousand passenger trips	Forecast (linear)	Deviation (linear)	Training-period approximation (MLP)	Single-year hold-out forecast (MLP)	Deviation (MLP)
2020	104 727	205 711	+96.4%	112 500	-	+7.4%
2021	138 209	202 987	+46.9%	135 100	-	-2.2%
2022	144 967	200 263	+38.1%	149 800	-	+3.3%
2023	240 403	197 539	-17.8%	-	236 900	-1.5%
Summary error metric	-	-	MAPE = 44.8%	In-sample MAPE = 4.3%	-	Hold-out APE = 1.5%

Note: Based on anonymized reports of Toshshahartransxizmat JSC. Linear 2020-2023 values are extrapolated from the 2014-2019 trend; MLP 2020-2022 values are in-sample approximations, and 2023 is the single hold-out forecast.

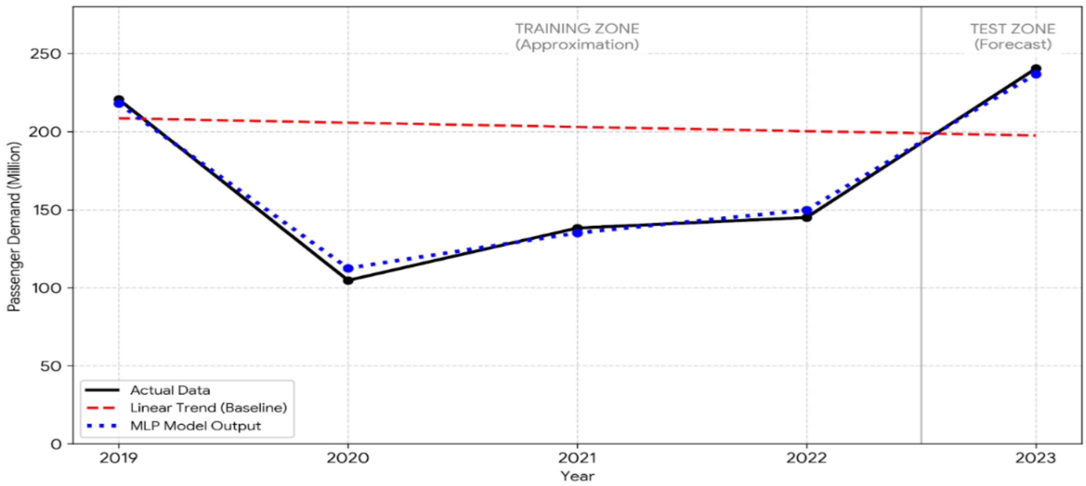

Figure 1: Conditional retrospective comparison for 2019-2023.

Table 4: Significant structural changes in operating indicators in 2019 and 2023.

Indicator	2019	2023	Change	Interpretation
Total recorded passenger trips, thousand passenger trips	220 577.9	240 403.0	+9.0%	Observed increase in recorded statistics
Electronically registered passenger trips, thousand passenger trips	116 650	195 199	+67.3%	Marker of increased digital trip capture in recorded statistics
Completed trips, units	4 176 286	7 994 834	+91.4%	Extensive growth in supply

Note: Table compiled by the author from anonymized operational reports of Toshshahartransxizmat JSC. Values for 2019 and 2023 are taken directly from these reports.

Table 5: Exogenously imposed annual $Y_{digital}$ inputs used in the 2024-2030 target scenario, including the observed 2023 anchor value.

Year	2023	2024	2025	2026	2027	2028	2029	2030
$Y_{digital}$ input, thousand passenger trips	195 199	259 224	326 592	383 025	442 981	497 737	545 735	583 884

Note: the 2023 value is observed and taken directly from anonymized operational reports of Toshshahartransxizmat JSC. Values for 2024-2030 are author-imposed exogenous scenario inputs rather than outputs of a separate forecasting model.

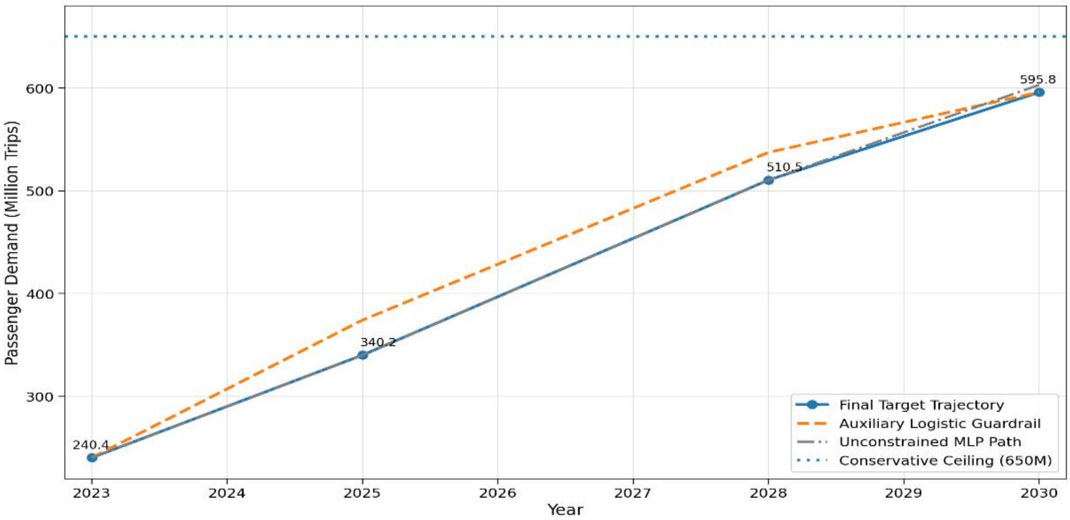

Figure 2: Scenario-based demand paths for 2024-2030. The solid blue line shows the final constrained trajectory, the dashed orange line shows the logistic guardrail, the dash-dotted gray line shows the unconstrained MLP path, and the dotted blue line marks the conservative ceiling (650 million passenger trips).

3.3 Scenario-Based Target Demand Indicators through 2030

The 2024-2030 path reported below is a scenario-conditioned planning trajectory rather than an autonomous forecast. For each year, the final value is defined as the pointwise minimum of the unconstrained MLP output and the auxiliary logistic guardrail specified in Section 2.4. Under this construction, the MLP supplies a non-linear conditional response pattern, whereas the logistic component serves only as an externally calibrated normative upper-bound path.

$Y_{digital}$ was not forecast endogenously. Instead, the 2024-2030 values were imposed exogenously as a monotonic saturation scenario anchored at the observed 2023 value, with larger annual increments in 2024-2027 and smaller ones in 2028-2030 as digital trip capture approached broader coverage.

The resulting 2030 trajectory should therefore be interpreted as a planning scenario conditioned by two external assumptions: exogenously imposed Ydigital inputs and a calibrated logistic guardrail. It is informative for scenario-based capacity planning, but it is not a purely endogenous model prediction.

The constrained planning trajectory suggests non-linear growth, with higher intensity in 2024-2027 and subsequent stabilization: by 2025, demand is projected to reach 340.2 million passenger trips (+41.5% relative to 2023); by 2028, the indicator is expected to reach 510.5 million; and by 2030, the trajectory approaches stabilization at 595.8 million passenger trips.

Under the adopted scenario assumptions, the model implies a projected increase in demand of approximately 2.5 times relative to 2023 by 2030. The resulting trajectory in Figure 2 is consistent with the target benchmarks of the "Uzbekistan-2030" strategy, which declares a multiple increase in passenger volumes via a modal shift from private vehicles to public transport [21].

The gray path is shown only to illustrate where the external guardrail becomes binding relative to the unconstrained non-linear scenario response.

4 DISCUSSION

The closer conditional fit of the MLP in the single 2023 hold-out year is consistent with its ability to represent non-linear changes during post-pandemic recovery and digital expansion. However, this result should not be interpreted as proof of stable forecasting superiority. In the present study, the neural-network component is used more modestly as a flexible element of a scenario-support framework for structurally disrupted annual data. The 67.3% rise in electronically registered passenger trips is treated solely as a descriptive marker of increased statistical visibility; no single causal mechanism is claimed. The findings should therefore be read as planning-oriented evidence from a small annual dataset rather than as structural identification of demand drivers.

To illustrate planning uncertainty rather than a single deterministic endpoint, two auxiliary $Y_{digital}$ trajectories were examined: slower saturation (-15% annual increments) and faster saturation (+15% annual increments). Under these scenario calculations, 2030 demand ranges from 562 to 629 million passenger trips. The central path of 595.8 million should therefore be interpreted only as a reference case within a broader planning corridor rather than as a point prediction.

The projected 2030 demand of 595.8 million passenger trips under the adopted scenario requires assessment of resource feasibility. According to internal reports of Toshshahartransxizmat JSC as of 01.06.2025, the average fleet size was 1916 units. Rolling-stock requirements were estimated under two scenarios.

If current productivity is maintained at 125 thousand passenger trips per vehicle per year, the rolling-stock deficit by 2030 will be 4766 - 1916 = 2850 units, representing a pessimistic upper bound. Eliminating this deficit solely by more than doubling the fleet would impose a critical burden on the street and road network [22].

This scenario reflects the potential effect of operational and managerial optimization under the assumed utilization improvement. If the fleet utilization coefficient rises from 0.706 to the target 0.85, a benchmark for efficient systems, productivity increases to 155 thousand passenger trips per vehicle per year. The required rolling stock under the improved-utilization scenario is then:

$$N_{req}^{opt} = \frac{Q_{2030}}{P_{bus}^{opt}} = \frac{595\,800\,000}{155\,000} \approx 3844, \quad (8)$$

Where:

- N_{req}^{opt} is the required rolling stock under the improved-utilization scenario;
- Q_{2030} is the projected annual passenger demand by 2030;
- P_{bus}^{opt} is the assumed annual vehicle productivity under the target utilization scenario, measured in passenger trips per vehicle per year.

According to (8), higher operational efficiency reduces procurement needs relative to the static case. The rolling-stock deficit by 2030 would therefore be 3844 - 1916 = 1928 units. Thus, process optimization through higher utilization lowers fleet procurement requirements by approximately 922 units, from 2850 to 1928, indicating potential economic efficiency of the proposed strategy under the stated assumptions.

The actual procurement need may be lower if planned infrastructure measures are implemented. Tashkent's development strategy to 2030 includes higher operating speeds through smart intersections and dedicated lanes, as well as metro expansion [23], [24]. These measures can raise effective carrying capacity and shift part of projected demand from surface transport to rail. Under this assumption, the projected surface-transport deficit may decline from 1928 units to about 1000. This estimate remains conditional on the timing and completeness of policy implementation [25].

From a managerial perspective, the proposed framework can be embedded in an annual decision-support cycle for fleet planning. At the end of each reporting period, the database is updated with new operational indicators, the conditional demand path is recalibrated, and the required fleet size is recomputed under alternative utilization scenarios. The resulting deficit estimate can then justify procurement volumes, route-capacity adjustments, and depot-level targets for the fleet utilization coefficient. In this form, the framework serves as an instrument of evidence-based transport-capacity planning and annual scenario support.

5 CONCLUSIONS

This study proposes a scenario-based planning framework for urban public transport demand under structural change and digital transformation in Tashkent. Within this framework, the MLP component is used not to support strong autonomous forecasting claims, but to provide a flexible conditional mapping within a broader planning structure that also includes exogenously specified $Y_{digital}$ inputs and an externally calibrated logistic guardrail. The proposed approach achieved an in-sample MAPE of 4.3% (2020-2022) and a 2023 hold-out APE of 1.5%; under the adopted scenario, passenger traffic is projected to reach 595.8 million passenger trips by 2030. The single 2023 hold-out result is therefore interpreted only as a limited case-specific check showing that the framework can produce planning-relevant conditional trajectories under the observed post-shock configuration, not as proof of stable generalization or robust comparative forecasting superiority.

Under the adopted scenario assumptions, the analysis indicates that accommodating the conditional 2030 demand trajectory would require not only fleet expansion but also operational-efficiency improvements, since higher fleet utilization and operating speed reduce procurement requirements.

Several limitations should be emphasized. First, the annual sample size, including only 9 training observations, severely restricts model complexity and the strength of generalization claims; the MLP component should therefore be read as a flexible heuristic approximator within a scenario-support framework rather than as a fully identified forecasting model. Second, out-of-sample evaluation is limited to a single hold-out year. Third, the 2030 trajectory is a hybrid scenario conditional on exogenous $Y_{digital}$ inputs and an externally calibrated normative logistic guardrail rather than a purely endogenous forecast. Fourth, the model produces point forecasts only, whereas investment planning would benefit from probabilistic intervals. Fifth, long-horizon results remain conditional on the timing and completeness of planned infrastructure measures; recalibration with new data is therefore essential. Within these limits, the framework may be used as a transparent scenario-support tool for transport-capacity planning and annual managerial recalibration, rather than as a stand-alone long-horizon forecasting model.

Ethics statement. The study uses anonymized aggregated data; no personal data or human-subject intervention were involved. Ethical approval was therefore not required.

Conflict of interest. The authors declare no conflict of interest.

Data availability. The raw data cannot be publicly released because they are based on anonymized internal operational reports. The structure of the analytical variables, the main data origins, the backward linear extrapolation rule used for early-year $N_{bus(t-1)}$ inputs, and the imposed $Y_{digital}$ scenario values for 2024-2030 are reported directly in the manuscript.

6 REFERENCES

[1] S. Patnaik, Ed., New Paradigm of Industry 4.0: Internet of Things, Big Data & Cyber Physical Systems, Studies in Big Data, vol. 64. Cham, Switzerland: Springer, 2020, [Online]. Available: https://doi.org/10.1007/978-3-030-25778-1.

[2] Y. Matseliukh, V. Lytvyn, and M. Bublyk, "Development of a neural network for forecasting passenger flows in smart city public electric transport," Technology Audit and Production Reserves, no. 5/2(85), pp. 20-25, 2025, [Online]. Available: https://doi.org/10.15587/2706-5448.2025.339550.

[3] R. A. Ataniyazova and L. A. Bozarov, "Analysis of the dynamics of demographic indicators of the population and housing construction of Tashkent by administrative districts" [in Russian], Zhurnal gumanitarnykh i estestvennykh nauk (Journal of Humanities and Natural Sciences), vol. 2, no. 4, pt. 2, pp. 188-196, Nov. 2023.

[4] World Bank, Public Transport Improvement Roadmap for the Tashkent Region. Washington, DC, USA: World Bank, Jan. 5, 2026, [Online]. Available: https://hdl.handle.net/10986/44119, [Accessed: Jan. 24, 2026].

[5] J. A. Abdazov, P. R. Kurbonov, N. Kh. Jaloliddinov, and R. Makhamadaliev, "Urbanization and housing infrastructure in Tashkent: an economic-geographic analysis," InterCarto. InterGIS, no. 31, pp. 638-653, 2025, [Online]. Available: https://doi.org/10.35595/2414-9179-2025-1-31-638-653.

[6] "Urban development of Uzbekistan," Strategy.uz, Oct. 22, 2025, [Online]. Available: https://strategy.uz/index.php?news=2178&lang=en, [Accessed: Jan. 24, 2026].

[7] B. O. Kenjaeva and A. A. Nazarov, "Logistics system of passenger transportation in large cities: problems and solutions," in S. M. Sultanova, N. U. Babakhanova, S. S. Shaumarov, and M. N. Masharipov, Eds., Sustainable Development of Transport, Lecture Notes in Networks and Systems, vol. 1366, ch. 7. Cham, Switzerland: Springer, 2025, [Online]. Available: https://doi.org/10.1007/978-3-031-88846-5_7.

[8] "Hisobotlar (Reports)" [in Uzbek], Toshshahartransxizmat JSC (Tashbus), n.d., [Online]. Available: https://tashbus.uz/news/#/hisobotlar, [Accessed: Jan. 24, 2026].

[9] M. Irisbekova, V. Nazarova, and M. Turgunov, "Methodology for using econometric models while transforming the market of transport and logistics services," in AIP Conference Proceedings, vol. 2476, no. 1, Art. no. 040006, May 2023, [Online]. Available: https://doi.org/10.1063/5.0107537.

[10] Gazeta.uz, "Bank cards equated to transport cards when paying fares on Tashkent public transport" [in Russian], Oct. 31, 2023, [Online]. Available: https://www.gazeta.uz/ru/2023/10/31/cards/, [Accessed: Jan. 24, 2026].

[11] M. Marupov, A. Murodov, Z. Yusufkhonov, and I. Absattorov, "Choice of optimum forecast models in planning and management of transport," AIP Conference Proceedings, vol. 3045, no. 1, Art. no. 050019, Mar. 2024, [Online]. Available: https://doi.org/10.1063/5.0197538.

[12] A. Utku and S. K. Kaya, "Multi-layer perceptron based transfer passenger flow prediction in Istanbul transportation system," Decision Making: Applications in Management and Engineering, vol. 5, no. 1, pp. 208-224, 2022, [Online]. Available: https://doi.org/10.31181/dmame0315052022u.

[13] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: generalization gap and sharp minima," arXiv preprint arXiv:1609.04836, 2016.

[14] M. Mohammadagha, S. Asadi, and H. Kazemi Naeini, "Evaluating machine learning performance using python for neural network models in urban transportation in New York city case study," Journal of Economy and Technology, vol. 4, pp. 266-283, 2026, [Online]. Available: https://doi.org/10.1016/j.ject.2025.11.001.

[15] I. C. Dipto, M. A. Rahman, T. Islam, and H. M. M. Rahman, "Prediction of accident severity using artificial neural network: a comparison of analytical capabilities between Python and R," Journal of Data Analysis and Information Processing, vol. 8, no. 3, pp. 134-157, Aug. 2020, [Online]. Available: https://doi.org/10.4236/jdaip.2020.83008.

[16] J. E. Leal and V. Parada, "Predicting modal choice for urban transport using an algebraic equation," Transportation Research Interdisciplinary Perspectives, vol. 22, Art. no. 100947, Nov. 2023, [Online]. Available: https://doi.org/10.1016/j.trip.2023.100947.

[17] B. Abdullaev, D. Yuldoshev, T. Muminov, and D. Axmedov, "Improving the method of assessing road safety at intersections of single-level highways," E3S Web of Conferences, vol. 264, Art. no. 05027, 2021, [Online]. Available: https://doi.org/10.1051/e3sconf/202126405027.

[18] M. M. Marupov, B. Aracıoğlu, and Z. Yusufkhon-Ugli Yusufkhonov, "Modeling and forecasting of freight traffic volumes," AIP Conference Proceedings, vol. 3045, no. 1, Art. no. 050028, Mar. 2024, [Online]. Available: https://doi.org/10.1063/5.0197412.

[19] CMWP Uzbekistan, IQ City Planning: Tashkent, 4Q 2023 [in Russian], 2023, [Online]. Available: https://cmwp.uz/images/IQ%20CITY%20PLANNING%20TASHKENT_CMWP_Q42023.pdf, [Accessed: Jan. 25, 2026].

[20] United Nations Economic Commission for Europe (UNECE), Sustainable Urban Mobility and Public Transport in UNECE Capitals. Geneva, Switzerland: United Nations, 2016.

[21] President of the Republic of Uzbekistan, Decree No. UP-158, Sep. 11, 2023, "On the Strategy 'Uzbekistan-2030'" [in Uzbek and Russian].

[22] A. Loder, L. Ambühl, M. Menendez, and K. W. Axhausen, "Understanding traffic capacity of urban networks," Scientific Reports, vol. 9, no. 1, Art. no. 16283, Nov. 8, 2019, [Online]. Available: https://doi.org/10.1038/s41598-019-51539-5.

[23] President of the Republic of Uzbekistan, Resolution No. PQ-368, Dec. 4, 2025, "On additional measures aimed at improving management of the transport system in Tashkent city and preventing traffic congestion" [in Russian].

[24] President of the Republic of Uzbekistan, Resolution No. PP-5260, Oct. 16, 2021, "On measures to improve the efficiency of the Tashkent Metro" [in Russian].

[25] Cabinet of Ministers of the Republic of Uzbekistan, Resolution No. 48, Jan. 18, 2019, "On approval of the Concept for implementing 'Smart City' technologies in the Republic of Uzbekistan" [in Russian].